

Springer Series in Reliability Engineering

Mircea Grigoriu

# Stochastic Systems

Uncertainty Quantification and  
Propagation

 Springer

# Springer Series in Reliability Engineering

*Series Editor*

Hoang Pham

For further volumes:

<http://www.springer.com/series/6917>

Mircea Grigoriu

# Stochastic Systems

Uncertainty Quantification and Propagation

Mircea Grigoriu  
Cornell University  
Hollister Hall  
14853 Ithaca  
NY, USA

ISSN 1614-7839

ISBN 978-1-4471-2326-2

ISBN 978-1-4471-2327-9 (eBook)

DOI 10.1007/978-1-4471-2327-9

Springer London Heidelberg New York Dordrecht

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2011941678

© Springer-Verlag London Limited 2012

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licenses issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))



*To Betsy*

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Introduction	1
1.2	Organization	5
1.3	Classroom Use	6
	References	7
<b>2</b>	<b>Essentials of Probability Theory</b>	<b>9</b>
2.1	Introduction	9
2.2	Probability Space	9
2.2.1	Sample Space	9
2.2.2	$\sigma$ -Field	10
2.2.3	Probability Measure	11
2.2.4	Construction of Probability Spaces	14
2.3	Measurable Functions and Random Elements	16
2.4	Independence	18
2.5	Sequence of Events	21
2.6	Expectation	22
2.7	Convergence of Sequences of Random Variables	27
2.8	Radon–Nikodym Derivative	31
2.9	Distribution and Density Functions	32
2.10	Characteristic Function	36
2.11	Conditional Expectation	38
2.12	Discrete Time Martingales	42
2.13	Monte Carlo Simulation	47
2.13.1	Gaussian Variables	48
2.13.2	Non-Gaussian Variables	48
2.13.3	Estimators	50
2.14	Exercises	53
	References	57

<b>3</b>	<b>Random Functions</b>	59
3.1	Introduction	59
3.2	Finite Dimensional Distributions	62
3.3	Sample Properties	64
3.4	Second Moment Properties	67
3.5	Weakly Stationary Random Functions	70
3.5.1	$\mathbb{R}$ -Valued Stochastic Processes	70
3.5.2	$\mathbb{R}^d$ -Valued Stochastic Processes	72
3.5.3	$\mathbb{R}$ -Valued Random Fields	73
3.6	Second Moment Calculus	76
3.6.1	Continuity	76
3.6.2	Differentiability	77
3.6.3	Integration	79
3.6.4	Spectral Representation	81
3.6.5	Karhunen-Loève Expansion	83
3.7	Classes of Stochastic Processes	85
3.7.1	Gaussian Random Functions	86
3.7.2	Translation Random Functions	86
3.7.3	Ergodic Random Functions	88
3.7.4	Markov Random Functions	90
3.7.5	Processes with Independent Increments	93
3.7.6	Continuous Time Martingales	95
3.8	Monte Carlo Simulation	109
3.8.1	Stationary Gaussian Random Functions	109
3.8.2	Translation Vector Processes	113
3.8.3	Non-Stationary Gaussian Processes	119
3.9	Exercises	124
	References	126
<b>4</b>	<b>Stochastic Integrals</b>	129
4.1	Introduction	129
4.2	Riemann-Stieltjes Integrals	130
4.3	Stochastic Integrals $\int BdB$ and $\int NdN$	131
4.4	Stochastic Integrals with Brownian Motion Integrators	136
4.4.1	Integrands in $\mathcal{H}_0^2$	137
4.4.2	Integrands in $\mathcal{H}^2$	138
4.4.3	Integrands in $\mathcal{H}$	141
4.5	Stochastic Integrals with Martingale Integrators	142
4.6	Stochastic Integrals with Semimartingale Integrators	146
4.7	Quadratic Variation and Covariation Processes	148
4.8	Exercises	152
	References	154

<b>5</b>	<b>Itô's Formula and Applications</b>	155
5.1	Introduction	155
5.2	Itô's Formula for $\mathbb{R}$ -Valued Semimartingales	155
5.2.1	Continuous Semimartingales	156
5.2.2	Arbitrary Semimartingales	158
5.3	Itô's Formula for $\mathbb{R}^d$ -Valued Semimartingales	162
5.4	Itô and Stratonovich Integrals	164
5.5	Applications	165
5.5.1	Stochastic Differential Equations	166
5.5.2	Tanaka's Formula	181
5.5.3	Random Walk Method	184
5.5.4	Girsanov's Theorem	193
5.6	Exercises	197
	References	198
<b>6</b>	<b>Probabilistic Models</b>	201
6.1	Introduction	201
6.2	Random Variables	203
6.2.1	Gaussian Variables	203
6.2.2	Translation Variables	206
6.2.3	Bounded Variables	210
6.2.4	Directional Wind Speed for Hurricanes	211
6.3	Random Functions	214
6.3.1	Systems with Uncertain Parameters	215
6.3.2	Inclusions in Multi-Phase Materials	217
6.3.3	Probabilistic Models for Microstructures	221
6.4	Exercises	233
	References	235
<b>7</b>	<b>Stochastic Ordinary Differential and Difference Equations</b>	237
7.1	Introduction	237
7.2	Stochastic Equations with Deterministic Coefficients	239
7.2.1	Discrete Time Linear Systems	239
7.2.2	Continuous Time Linear Systems	241
7.2.3	Continuous Time Nonlinear Systems	248
7.3	Stochastic Difference Equations with Random Coefficients	254
7.3.1	General Considerations	255
7.3.2	Monte Carlo Simulation	260
7.3.3	Conditional Analysis	261
7.3.4	Stochastic Reduced Order Models	265
7.3.5	Stochastic Galerkin and Collocation Methods	267
7.3.6	Taylor Series	267
7.3.7	Perturbation Series	269

7.4	Stochastic Differential Equations with Random Coefficients . . .	271
7.4.1	General Considerations . . . . .	271
7.4.2	Monte Carlo Simulation . . . . .	276
7.4.3	Conditional Analysis . . . . .	277
7.4.4	Conditional Monte Carlo Simulation . . . . .	278
7.4.5	State Augmentation . . . . .	286
7.4.6	Stochastic Reduced Order Models . . . . .	289
7.4.7	Stochastic Galerkin Method. . . . .	290
7.4.8	Stochastic Collocation Method. . . . .	295
7.4.9	Taylor, Perturbation, and Neumann Series . . . . .	300
7.5	Applications . . . . .	303
7.5.1	Stochastic Stability. . . . .	304
7.5.2	Noise Induced Transitions. . . . .	311
7.5.3	Solution of Uncertain Dynamic Systems by SROMs. . .	317
7.5.4	Degrading Systems. . . . .	327
7.6	Exercises. . . . .	331
	References . . . . .	333
<b>8</b>	<b>Stochastic Algebraic Equations. . . . .</b>	<b>337</b>
8.1	Introduction. . . . .	337
8.2	SAEs with Arbitrary Uncertainty . . . . .	338
8.2.1	General Considerations. . . . .	339
8.2.2	Monte Carlo Method . . . . .	342
8.2.3	Stochastic Reduced Order Model Method . . . . .	344
8.2.4	Stochastic Galerkin Method. . . . .	357
8.2.5	Stochastic Collocation Method. . . . .	363
8.2.6	Reliability Method . . . . .	367
8.3	SAEs with Small Uncertainty . . . . .	368
8.3.1	Taylor Series. . . . .	369
8.3.2	Perturbation Series . . . . .	371
8.3.3	Neumann Series. . . . .	373
8.3.4	Equivalent Linearization . . . . .	374
8.4	Exercises. . . . .	375
	References . . . . .	377
<b>9</b>	<b>Stochastic Partial Differential Equations. . . . .</b>	<b>379</b>
9.1	Introduction. . . . .	379
9.2	Stochastic Partial Differential Equations . . . . .	380
9.3	Discrete Approximations of SPDEs . . . . .	386
9.4	Applied SPDEs: Arbitrary Uncertainty . . . . .	393
9.4.1	General Considerations . . . . .	394
9.4.2	Deterministic Boundary Value Problems. . . . .	396
9.4.3	Stochastic Boundary Value Problems . . . . .	398

9.4.4	Monte Carlo Simulation . . . . .	403
9.4.5	Stochastic Reduced Order Model Method . . . . .	414
9.4.6	Stochastic Reduced Order Models . . . . .	414
9.4.7	Extended Stochastic Reduced Order Models . . . . .	427
9.4.8	Stochastic Galerkin Method. . . . .	431
9.4.9	Stochastic Collocation Method. . . . .	440
9.5	Applied SPDEs: Small Uncertainty . . . . .	446
9.5.1	Taylor Series . . . . .	447
9.5.2	Perturbation Series . . . . .	448
9.5.3	Neumann Series . . . . .	449
9.6	Exercises. . . . .	451
	References . . . . .	452
<b>Appendix A: Parametric Models, Quantizers, and Stochastic Reduced Order Models . . . . .</b>		<b>455</b>
<b>Appendix B: A Primer on Functional Analysis . . . . .</b>		<b>475</b>
<b>Index . . . . .</b>		<b>519</b>

# Chapter 1

## Introduction

### 1.1 Introduction

Equations are used in applied sciences and engineering to characterize the state of mechanical, economics, physical, environmental, and other systems. For a long time it has been assumed that the coefficients of, the input to, and the end conditions for these equations are deterministic and perfectly known. Yet, owing to inherent variability and/or incomplete knowledge, the entries of most state equations are uncertain. Equations with deterministic, perfectly known entries are called deterministic equations (DEs); those with random entries are called stochastic equations (SEs). Although DEs are useful in many cases, they may provide limited information even on the trend of a system state. In contrast, SEs can capture both the trend and variability of a system state.

The study of SEs expanded significantly during the last decades in both mathematical and applied literatures. The class of stochastic equations in the mathematical literature is a rather small subset of that encountered in applications. The thrust of most mathematical studies is on the establishment of conditions for the existence and uniqueness of the solution of SEs [1–7]. On the other hand, the focus of most applied studies is the development of accurate and efficient methods for calculating solution statistics. Technical details related to the existence and uniqueness of the solution of these equations are frequently overlooked.

The main objective of the book is to foster interactions between engineers, scientists, and mathematicians with a view to promote the development of accurate and efficient methods for solving SEs based on rigorous mathematical arguments. It is hoped that these interactions will introduce engineers and scientists to rigorous mathematical arguments and apprise mathematicians of novel theoretical problems rooted in applications.

Following are examples of DEs encountered in various applied fields together with their stochastic counterparts. Potential limitations of DEs are illustrated by a simple mechanical system. Let  $U(x)$  denote the displacement function for a beam with length  $l > 0$  fixed at the left end and free at the right end, a simple mechanical

system. This function satisfies at equilibrium the ordinary differential equation

$$K(x) \frac{d^2 U(x)}{dx^2} = -M(x), \quad x \in (0, l), \quad (1.1)$$

with boundary conditions  $U(0) = 0$  and  $U'(0) = 0$ , where  $K(x)$  is the beam stiffness,  $M(x) = -\int_x^l (\xi - x) Q(\xi) d\xi$  denotes the bending moment, and  $Q(x)$  is the action on the beam. The price  $X(t)$  of a stock at time  $t$  and its evolution in time, a commonly used descriptor in equity market, can be modeled by an ordinary differential equation of the type

$$\frac{dX(t)}{dt} = cX(t) + \sigma X(t) (\text{"noise"})(t), \quad t \geq 0, \quad (1.2)$$

where the constants  $c$  and  $\sigma$  are mean return rate and volatility, respectively, and  $(\text{"noise"})(t)$  denotes a function of time capturing market fluctuations ([6], Chap. 10). The pollutant concentration  $U(x, t)$  at location  $x \in D$  and time  $t$  in a medium with permeability  $\Sigma(x)$  under a flux  $W(x, t)$ , a metric of great environmental interest, satisfies the partial differential equation

$$\frac{\partial U(x, t)}{\partial t} = \nabla \cdot (\Sigma(x) \nabla U(x, t)) + W(x, t), \quad x \in D, \quad t \geq 0, \quad (1.3)$$

accompanied by initial and boundary conditions, where  $\nabla$  denotes the differential operator  $(\partial/\partial x_1, \dots, \partial/\partial x_d)$  and  $D$  is bounded subset in  $\mathbb{R}^d$ ,  $d \leq 3$ .

Analytical solutions for differential equations are only possible in simple cases. In most applications, these equations need to be solved by finite difference, finite element, or other numerical methods that involve space and/or time discretization. The equations generated by numerical methods differ from the original differential equations. For example, the finite difference approximation of (1.1) is an algebraic equation of the form

$$AV + C = 0, \quad (1.4)$$

where  $V$  is a vector collecting values of  $U(x)$  at a finite number of points in the domain of definition of this equation,  $A$  is a square matrix whose entries depend on beam stiffness and boundary conditions, and  $C$  collects values of  $M(x)$  at a finite number of spatial coordinates. The discrete version of (1.2) is a recurrence formula giving future stock values as a function of their current values and market volatility. The finite difference approximation of the partial differential equation in (1.3) is the ordinary differential equation

$$\frac{dV(t)}{dt} = AV(t) + C(t), \quad (1.5)$$

where  $A$  depends on medium permeability  $\Sigma(x)$  and boundary conditions, and the coordinates of  $V(t)$  and  $C(t)$  are values of  $U(x, t)$  and  $W(x, t)$  at a finite number of spatial coordinates.



A common assumption in (1.1)–(1.3) is that both the coefficients of and the input to these equations are deterministic and perfectly known. In this classical formulation,  $K(x)$  and  $Q(x)$  in (1.1);  $c$ ,  $\sigma$ , (“noise”)( $t$ ), and initial state  $X(0)$  in (1.2); and  $\Sigma(x)$ ,  $W(x, t)$ , and the associated initial and boundary conditions in (1.3) are perfectly known, deterministic functions, so that the state of these systems is described by solutions of deterministic equations. The theory of DEs is well established and is discussed in numerous books [8–13].

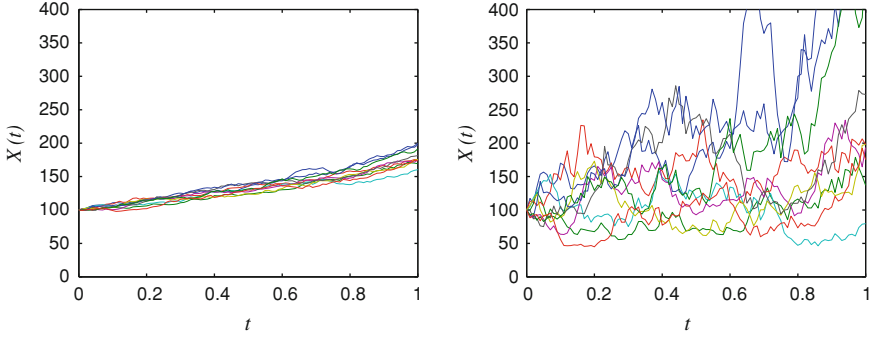
As previously stated, the classical formulation using DEs to describe problems can be unsatisfactory since (i) the coefficients, the input, and/or the end conditions of these equations can be uncertain due to incomplete information, limited understanding of some phenomena, complex relationships between microscopic and macroscopic properties of materials, and inherent randomness and (ii) DEs deliver a single solution for a system state, rather than a range of likely solutions needed to characterize the performance of systems encountered in applications, that is, systems with uncertain properties. For example, suppose stiffness  $K(x) = K$  and action  $Q(x) = Q$  in (1.1) are space invariant, so that  $U(l) = Ql^4/(8K)$ . If  $K$  and  $Q$  are uncertain parameters described by independent, uniformly distributed random variables in the ranges  $(k_1, k_2)$ ,  $0 < k_1 < k_2$  and  $(q_1, q_2)$ ,  $0 < q_1 < q_2$ , respectively, the first two moments of the tip displacement  $U(l)$  are

$$\begin{aligned} E[U(l)] &= \frac{l^4}{8} \frac{q_1 + q_2}{2} \frac{\ln(k_2/k_1)}{k_2 - k_1} \\ E[U(l)^2] &= \left(\frac{l^4}{8}\right)^2 \frac{q_1^2 + q_2^2 + q_1 q_2}{3k_1 k_2}. \end{aligned} \quad (1.6)$$

The tip displacement corresponding to the average values,  $\bar{q} = (q_1 + q_2)/2$  and  $\bar{k} = (k_1 + k_2)/2$ , of  $Q$  and  $K$ , that is, the solution of a DE with  $(Q, K)$  set equal to  $(\bar{q}, \bar{k})$  is  $\bar{u}(l) = \bar{q}l^4/(8\bar{k})$ . For  $(q_1, q_2) = (1, 2)$  and  $(k_1, k_2) = (0.5, 1)$ , we have  $E[U(l)] = 2.0794(l^4/8)$ ,  $\text{Std}[U(l)] = (E[U(l)^2] - E[U(l)]^2)^{1/2} = 0.5853(l^4/8)$ , and  $\bar{u}(l) = 2(l^4/8)$ . The solution of DE approximates satisfactorily the mean  $E[U(l)]$  of  $U(l)$ , but provides no information on the likely range of values of the tip displacement. However, even the mean of  $U(l)$  may be missed by DEs. For example,  $E[U(l)] = 3.8376(l^4/8)$  and  $\bar{u}(l) = 2.7273(l^4/8)$  for  $Q$  as previously and  $K$  uniformly distributed in  $(0.1, 1)$ .

Following are the stochastic counterparts of the deterministic equations in (1.1)–(1.5). The equation (1.1) becomes a stochastic ordinary differential equation if action  $Q(x)$  and/or stiffness  $K(x)$  are random functions, and so does (1.2) if  $(c, \sigma)$  are random variables and/or (“noise”)( $t$ ) is a stochastic process; (1.3) becomes a stochastic partial differential equation if permeability  $K(x)$ , flux  $W(x, t)$ , and/or initial and boundary conditions are random functions; (1.4) becomes a stochastic algebraic equation if it corresponds to the stochastic version of (1.1); and (1.5) becomes a stochastic ordinary differential equation if the permeability  $\Sigma(x)$  and the flux  $W(x, t)$  are random functions.

The white noise input model considered almost exclusively in mathematical studies precludes the use of classical calculus since some of the integrals in the



**Fig. 1.1** Samples of stock prices for  $(c, \sigma) = (0.5, 0.1)$  (left panel) and  $(c, \sigma) = (0.5, 1)$  (right panel)

expression of the solution are not defined in the Riemann–Stieltjes sense. For example, the noise in (1.2) is usually viewed as the formal derivative of a standard Brownian motion  $B(t)$ ,  $t \geq 0$ , a process with independent Gaussian increments, mean 0, variance  $E[B(t)^2] = t$ , and continuous, non-differentiable samples. The corresponding differential and integral forms of (1.2) with “(noise)( $t$ )” replaced with the formal derivative  $dB(t)/dt$  of a Brownian motion are

$$\begin{aligned} dX(t) &= cX(t)dt + \sigma X(t)dB(t), \quad t \geq 0, \quad \text{and} \\ X(t) &= x_0 + c \int_0^t X(s)ds + \sigma \int_0^t X(s)dB(s), \quad t \geq 0, \end{aligned} \quad (1.7)$$

for  $X(0) = x_0$ . The integral  $\int_0^t X(s)dB(s)$  in (1.7) is an Itô rather than Riemann–Stieltjes integral. The solution of the Itô stochastic equations in (1.7) is

$$X(t) = x_0 \exp \left[ (c - \sigma^2/2)t + \sigma B(t) \right], \quad (1.8)$$

and can be obtained by Itô’s calculus ([14], Example 1.3.8). The plots in the left and the right panels of Fig. 1.1 show samples of  $X(t)$  for  $(c, \sigma) = (0.5, 0.1)$  and  $(c, \sigma) = (0.5, 1)$ . The significant difference between samples of  $X(t)$  corresponding to the same mean return rate  $c$  and noise  $B(t)$ , but different volatilities, shows that the use of a single (deterministic) value for  $\sigma$  is unrealistic if  $\sigma$  is uncertain since  $X(t)$  is sensitive to the particular value used for this parameter.

The complexity of most SEs encountered in applications has prevented the development of general and rigorous methods for solving these equations. Most methods for solving SEs in the applied literature are based on heuristic arguments and approximations whose quality may be difficult to assess. Rigorous methods from the mathematical literature can be used for the class of SEs that can be recast in standard format, that is, stochastic equations with deterministic coefficients driven by white noise. For example, if the driving noise in (1.2) is a colored noise  $Z(t)$  defined by  $dZ(t) = -\rho Z(t)dt + \sqrt{2\rho}dB(t)$ ,  $\rho > 0$ ,  $t \geq 0$ , then the state  $(X, Z)$  satisfies the differential equation

$$\begin{cases} dX(t) = cX(t)dt + \sigma X(t)Z(t)dt \\ dZ(t) = -\rho Z(t)dt + \sqrt{2\rho}dB(t), \end{cases} \quad (1.9)$$

that is, an ordinary stochastic differential equation with deterministic coefficients driven by white noise. Similar arguments hold for discrete versions of stochastic equations of the type in (1.5) if  $C(t)$  can be represented as the output of a filter to white noise. It is also possible in some cases to convert SEs with random coefficients into SEs with deterministic coefficients. For example, if the volatility  $\sigma$  in (1.7) is uncertain and modeled by a random variable  $Y$ , it can be viewed as the solution of  $\dot{\sigma}(t) = 0$  with initial condition  $\sigma(0) = Y$ , so that the bivariate process  $(X(t), \sigma(t))$  satisfies an ordinary differential equation driven by white noise. Other examples in which this conversion is possible can be found in [15].

## 1.2 Organization

The book is largely self-contained with nine chapters and two appendices providing a summary of useful facts. [Chapters 2 and 3](#) review essential concepts on probability theory and random functions, present Monte Carlo algorithms for generating samples of random variables and functions, and construct estimators for properties of random elements. [Chapter 4](#) illustrates the need for alternatives to the Riemann–Stieltjes integral, defines the Itô integral, and establishes properties of these integrals. [Chapter 5](#) establishes Itô’s formula for continuous and arbitrary semimartingales, defines stochastic differential equations, and gives conditions for the existence and uniqueness of the solutions for these equations. The Itô formula is applied to develop moment equations for the state of dynamic systems driven by random noise, solve locally a class of deterministic partial differential equations, and establish Girsanov’s formula.

The focus of the remaining chapters is on methods for solving stochastic algebraic and differential equations. [Chapter 6](#) examines methods for constructing probabilistic models for the random entries of stochastic equations. Models for random variables and functions are used to describe random microstructures. Linear models with random coefficients used almost exclusively to characterize the coefficients of elliptic stochastic partial differential equations are also examined. [Chapter 7](#) considers ordinary differential equations with random input and deterministic and random coefficients. Itô’s formula is applied to develop Fokker–Planck and other equations for the state of dynamic deterministic systems driven by random noise, that are commonly used in the random vibration theory. Equations with random coefficients and input are solved by Monte Carlo simulation, conditional analysis, stochastic reduced order models, stochastic Galerkin, and stochastic collocation. Numerical examples are presented to illustrate the implementation of these methods and assess their performance. The special case of equations with coefficients of small uncertainty are solved by Taylor, Neumann, and perturbation series. [Chapter 8](#) deals with stochastic algebraic equations, and solves these equations by Monte Carlo, stochastic reduced order models, stochastic Galerkin, stochastic collocation, and reliability

methods. Numerical examples provide insight on the relative accuracy and efficiency of these methods. Taylor, Neumann, and perturbation series are applied to solve stochastic algebraic equations with random entries of small uncertainty. [Chapter 9](#) considers stochastic partial differential equations. It is first noted that the methods in [Chaps. 7–8](#) can be used to solve discrete versions of these equations. Then, an equation of the type considered in the mathematical literature is defined and solved. Following these preliminary considerations, equations commonly encountered in applications are defined and solved by Monte Carlo, stochastic reduced order models, stochastic Galerkin, and stochastic collocation. Numerical examples illustrate the application of these methods and their relative performance. The special case of these equations with random entries of small uncertainty are solved by Taylor, Neumann, and perturbation series. The presentation in [Chaps. 7–9](#) uses concepts introduced in [Chaps. 2–5](#) and Appendices A–B and probabilistic models constructed in [Chap. 6](#).

### 1.3 Classroom Use

This book can be used as text for three one-semester graduate courses emphasizing various topics from the book. Following are potential titles and contents of these courses.

- **Formulation of stochastic equations.**  
The course may review essentials of probability theory and random functions ([Chaps. 2 and 3](#)), discuss the construction of probabilistic models for both the coefficients of and the input to stochastic equations ([Chap. 6](#)), Monte Carlo algorithms for generating samples of random elements considered in [Chaps. 2 and 3](#), and Monte Carlo solutions of stochastic equations discussed in [Chaps. 7–9](#).
- **Random vibration by Itô’s calculus.**  
The course may largely follow developments in the first part of [Chap. 7](#) dealing with linear and nonlinear dynamic systems subjected to random noise. The derivations in this chapter use concepts in [Chaps. 4 and 5](#) on stochastic integrals, Itô’s formula, and stochastic differential equations.
- **Stochastic partial differential equations.**  
The study of stochastic elliptic partial differential equations and applications to random heterogeneous materials in [Chap. 9](#) provide ample material for this course. The course can be extended to a two semester course by including the class of stochastic equations examined in [Chaps. 7 and 8](#) and facts from Appendix B.

**Acknowledgments** This book could not have been completed without the contributions of many individuals. In particular, I wish to express my deepest appreciation to Professors S. I. Resnick and G. Samorodnitsky of Cornell University, Professor F. Potra of the University of Maryland Baltimore County, and Professor S. T. Ariaratnam of the University of Waterloo, Canada, for numerous technical discussions, Dr. E. Simiu of National Institute of Standards and Technology

for reviewing the entire manuscript, Professors W. Aquino and C. Earls for useful comments on [Chap. 9](#), my doctoral student S. Saroukhani for helping with some numerical examples, Mr. C. Willkens for his enthusiastic and professional support of the computer hardware and software used in this project, and my son Bogdan for his help with figures, index, and other aspects of the book. Finally, I am grateful to my wife Betsy for understanding, encouragement, support, as well as most valuable comments to this chapter.

A significant part of the work incorporated in this book has been supported by the National Science Foundation, Sandia National Laboratories, National Institute of Standards and Technology, Jet Propulsion Laboratory, Air Force Office of Scientific Research, Federal Aviation Administration, and other institutions. I am indebted to these organizations and their continued support.

## References

1. Chung KL, Williams RJ (1990) Introduction to stochastic integration. Birkhäuser, Boston
2. Dalang RC, Khoshnevisan D, Mueller C, Nualart D, Xiao Y (2006) A minicourse on stochastic partial differential equations. Springer, New York
3. Holden H, Øksendal B, Ubøe J, Zhang T (1996) Stochastic partial differential equations. Birkhäuser, Boston
4. Kuo H-H (2006) Introduction to stochastic integration. Springer, New York
5. Protter P (1990) Stochastic integration and differential equations. Springer, New York
6. Steele JM (2001) Stochastic calculus and financial applications. Springer, New York
7. Tsokos CP, Padgett WJ (1974) Random integral equations with applications in life sciences and engineering. Academic Press, New York
8. Boresi AP, Chong KP (2000) Elasticity in engineering mechanics. Wiley, New York
9. Cakmak AS, Botha JF, Gray WG (1987) Computational and applied mathematics for engineering analysis. Springer, New York
10. Friedman B (1969) Lectures on applications-oriented mathematics. Wiley, New York
11. Hairer E, Nørsett SP, Wanner G (1993) Solving ordinary differential equations I. Nonstiff problems, 2nd edn. Springer, New York
12. Hildebrand FB (1965) Methods of applied mathematics. Prentice Hall, Englewood Cliffs
13. John F (1982) Partial differential equations, 4th edn. Springer, New York
14. Mikosch T (1998) Elementary stochastic calculus. World Scientific, River Edge
15. Whittle P (1954) On stationary processes in the plane. *Biometrika* 41(3/4):434–449

# Chapter 2

## Essentials of Probability Theory

### 2.1 Introduction

Many properties of physical systems and/or actions of these systems are uncertain, so that the behavior of these systems cannot be forecasted in a precise deterministic manner; it can only be described probabilistically. For example, the weather person tells us the chance of rain tomorrow. Engineers calculate the likelihood that a particular mechanical system will perform according to specified standards.

Suppose a relevant output of a physical system depends on a finite number of uncertain parameters. For weather forecasting, these parameters relate to the current meteorologic conditions and atmospheric processes. For aircraft design, these parameters include material properties, state of electronic components, and flight patterns. Our objective is to calculate the probability that an output of interest has specified features, such as the chance of rain tomorrow for the weather person or the adequate aircraft performance for the aircraft engineer. This chapter provides the framework for calculating these types of probabilities.

### 2.2 Probability Space

A probability space  $(\Omega, \mathcal{F}, P)$  consists of a sample space  $\Omega$ , a  $\sigma$ -field  $\mathcal{F}$ , and a probability measure  $P$ . Sample spaces,  $\sigma$ -fields, and probability measures are defined and illustrated by examples.

#### 2.2.1 Sample Space

**Definition 2.1** A set  $\Omega$  collecting the outcomes of a particular experiment is called sample space.

For example,  $\Omega = \{\text{head, tail}\}$ ,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , and  $\Omega = [a, b] \subset [0, \infty)$  are sample spaces for the experiments of tossing a coin, rolling a dice, and measuring daily rainfall amounts in Ithaca, NY. The first two sample spaces and the last sample space have, respectively, a finite and an uncountable number of elements. Sample spaces can be finite, countable, or uncountable.

### 2.2.2 $\sigma$ -Field

Consider a game in which one wins \$10 and loses \$5 for outcomes of a die rolling experiment in  $\{1, 2\}$  and  $\{3, 4, 5, 6\}$ , respectively. The particular outcome  $\{\omega\}$  is not relevant. The relevant information is contained in  $\mathcal{A} = \{\{1, 2\}, \{3, 4, 5, 6\}\}$  since one wins \$10 if  $\omega \in \{1, 2\}$  and loses \$5 if  $\omega \in \{3, 4, 5, 6\}$ .  $\sigma$ -fields are subsets of  $\Omega$  that are relevant for a particular objective. The  $\sigma$ -field for the game considered here includes  $\mathcal{A}$ .

**Definition 2.2** A collection  $\mathcal{F}$  of subsets of  $\Omega$  is said to be a  $\sigma$ -field on a sample space  $\Omega$  if (1)  $\emptyset \in \mathcal{F}$ , (2)  $A \in \mathcal{F}$  implies  $A^c \in \mathcal{F}$ , and (3)  $A_i \in \mathcal{F}$ ,  $i \in I$ , implies  $\bigcup_{i \in I} A_i \in \mathcal{F}$ , where  $I$  is a countable set. The members of  $\mathcal{F}$  are called events, or  $\mathcal{F}$ -measurable subsets of  $\Omega$ , or just measurable subsets of  $\Omega$ . The pair  $(\Omega, \mathcal{F})$  is said to be a measurable space.

The first and the third conditions in the previous definition can be replaced with  $\Omega \in \mathcal{F}$  and  $\bigcap_{i \in I} A_i^c \in \mathcal{F}$  by using condition (2) and De Morgan's formula. Also note that the last two conditions in the definition of  $\mathcal{F}$  are consistent with our intuition, which suggests that  $A^c$  is observed if  $A$  is not and that  $\bigcup_{i \in I} A_i$  is observed if a subset of  $\{A_i, i \in I\}$  is.

*Example 2.1* The  $\sigma$ -field  $\mathcal{F}$  associated with the game in which one wins \$10 and loses \$5 for outcomes of a die rolling experiment in  $\{1, 2\}$  and  $\{3, 4, 5, 6\}$  is  $\mathcal{F} = \{\emptyset, \{1, 2\}, \{3, 4, 5, 6\}, \Omega\}$ , where  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . If the game is modified such that one wins \$10 if  $\omega \in \{1, 2\}$  and loses \$3, \$4, \$5, and \$6 for outcomes  $\omega = 3, 4, 5$ , and 6, respectively,  $\mathcal{F}$  is too coarse to capture all relevant events; it needs to be refined to include the events  $\{1, 2\}$ ,  $\{3\}$ ,  $\{4\}$ ,  $\{5\}$ , and  $\{6\}$ , and complements and unions of these events.  $\diamond$

*Example 2.2* Atoms are the finest members of a  $\sigma$ -field  $\mathcal{F}$ , that is,  $A \in \mathcal{F}$  is an atom of  $\mathcal{F}$  if any event  $B \in \mathcal{F}$  included in  $A$  is either  $\emptyset$  or  $A$ . The sets  $\{1, 2\}$  and  $\{3, 4, 5, 6\}$  are atoms of the first  $\sigma$ -field and the sets  $\{1, 2\}$ ,  $\{3\}$ ,  $\{4\}$ ,  $\{5\}$ ,  $\{6\}$  are atoms of the second  $\sigma$ -field (Example 2.1).  $\diamond$

**Definition 2.3** The  $\sigma$ -field generated by a collection of subsets  $\mathcal{A}$  of  $\Omega$  is

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{G} \supseteq \mathcal{A}} \mathcal{G}, \quad \text{where } \{\mathcal{G}\} \text{ are } \sigma\text{-fields on } \Omega. \quad (2.1)$$

There is no  $\sigma$ -field smaller than  $\sigma(\mathcal{A})$  that includes  $\mathcal{A}$ .

**Example 2.3** Let  $S$  be a metric space and let  $\mathcal{T}$  be the topology induced on  $S$  by its metric (Sect. B.1.1). The Borel  $\sigma$ -field on  $S$  is the  $\sigma$ -field  $\mathcal{S} = \sigma(\mathcal{T})$  generated by  $\mathcal{T}$ . The members of  $\sigma(\mathcal{T})$  are called Borel sets. The Borel  $\sigma$ -fields on  $S = \mathbb{R}^d$ ,  $d > 1$ , and  $S = \mathbb{R}$  are generated by the intervals in these spaces and are denoted by  $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}^d$  and  $\mathcal{B}(\mathbb{R}) = \mathcal{B}^1 = \mathcal{B}$ , respectively. The Borel sets on  $\mathbb{R}$  can be generated by open, semi-open, finite, semi-finite, and other intervals of the real line ([11], Sect. 1.7).  $\diamond$

### 2.2.3 Probability Measure

We define measures and probability measures, review properties of probability measures useful for applications, and use conditional probabilities to introduce the law of total probability and Bayes theorem.

**Definition 2.4** Let  $(\Omega, \mathcal{F})$  be a measurable space. A set function  $\mu : \mathcal{F} \rightarrow [0, \infty]$  such that (1)  $\mu(\emptyset) = 0$  and (2)  $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$  for mutually disjoint sets  $A_n \in \mathcal{F}$  is called a measure on  $(\Omega, \mathcal{F})$ . If  $(\Omega, \mathcal{F}) = (\mathbb{R}^d, \mathcal{B}^d)$ , then  $\mu$  is said to be a Borel measure. The triple  $(\Omega, \mathcal{F}, \mu)$  is called a measure space.

**Definition 2.5** If  $\mu$  has finite total mass, that is,  $\mu(\Omega) < \infty$ , the measure  $\mu$  is said to be finite. In this case, it can be normalized to take values in  $[0, 1]$ . Any measure with unit total mass is called a probability measure and is denoted by  $P$ . The triple  $(\Omega, \mathcal{F}, P)$  is called a probability space.

**Definition 2.6** A measure  $\mu$  on a measurable space  $(\Omega, \mathcal{F})$  is said to be  $\sigma$ -finite if there exists a countable, pairwise disjoint sequence of measurable sets  $\{A_n\}$ , that is,  $A_n \in \mathcal{F}$ ,  $n = 1, 2, \dots$ , such that  $\mu(A_n) < \infty$ ,  $n = 1, 2, \dots$ , and  $\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap A_n)$  for every  $A \in \mathcal{F}$  ([5], Proposition 13).

The Lebesgue measure  $\lambda$  on  $(\mathbb{R}, \mathcal{B})$  is  $\sigma$ -finite since there exists a pairwise disjoint sequence  $A_n = [n, n+1)$ ,  $n \in \mathbb{Z}$ , such that  $\lambda(A_n) < \infty$  and  $\lambda(A) = \sum_{n=1}^{\infty} \lambda(A \cap A_n)$  for every  $A \in \mathcal{B}$ . Similar arguments show that the Lebesgue measure on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  is  $\sigma$ -finite. However, the Lebesgue measure is not finite.

**Definition 2.7** Let  $(\Omega, \mathcal{F}, P)$  be a probability space. A set  $N \in \mathcal{F}$  such that  $P(N)=0$  is called a null set. A property valid on  $\Omega \setminus N$  is said to hold almost everywhere (a.e.), almost surely (a.s.), for almost every  $\omega$ , or with probability one (w.p.1.).

**Definition 2.8** A probability space  $(\Omega, \mathcal{F}, P)$  is complete if  $A \subset B$  such that  $B \in \mathcal{F}$  and  $P(B) = 0$ , then  $A \in \mathcal{F}$ , which implies  $P(A) = 0$ , by the second condition in Definition 2.2.

It is assumed throughout the book that the probability spaces are complete. The assumption is not restrictive since for any probability space  $(\Omega, \mathcal{F}, P)$  there exists a complete space  $(\Omega, \overline{\mathcal{F}}, \overline{P})$  such that  $\mathcal{F} \subseteq \overline{\mathcal{F}}$  and  $P = \overline{P}$  on  $\mathcal{F}$  ([4], Theorem



2.2.5). The completion of measures is rather simple ([1], p. 4). For example, set  $\mathcal{N} = \{B \subseteq \Omega : \exists N \in \mathcal{F} \text{ with } P(N) = 0 \text{ and } B \subset N\}$ ,  $\overline{\mathcal{F}} = \{A \cup B : A \in \mathcal{F}, B \in \mathcal{N}\}$ , and  $\overline{P}(A \cup B) = P(A)$ , where  $A \in \mathcal{F}$  and  $B \in \mathcal{N}$ . Then  $\overline{P}$  is a probability measure on  $(\Omega, \overline{\mathcal{F}})$  and  $\overline{\mathcal{F}}$  is a  $\sigma$ -field.

The following properties of the probability measure  $P$  are useful for calculations, and result directly from its definition.

$$\begin{aligned}
 P(A) &= 1 - P(A^c), \quad \text{for } A \in \mathcal{F}, \\
 P(A) &\leq P(B), \quad \text{for } A \subseteq B, \quad A, B \in \mathcal{F}, \\
 P(\cup_{i=1}^{\infty} A_i) &\leq \sum_{i=1}^{\infty} P(A_i), \quad \text{for } A_i \in \mathcal{F}, \\
 P(A \cup B) &= P(A) + P(B) - P(A \cap B), \quad \text{for } A, B \in \mathcal{F}, \text{ and} \\
 P(B) &= \sum_{i=1}^n P(B \cap A_i), \quad \text{for } A_i, B \in \mathcal{F} \text{ and } A_1, \dots, A_n \text{ a partition of } \Omega.
 \end{aligned} \tag{2.2}$$

The probability measure also satisfies the inclusion–exclusion formula

$$\begin{aligned}
 P(\cup_{i=1}^n A_i) &= \sum_{i=1}^n P(A_i) - \sum_{i=2}^n \sum_{j=1}^{i-1} P(A_i \cap A_j) \\
 &\quad + \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(\cap_{q=1}^n A_q).
 \end{aligned} \tag{2.3}$$

**Theorem 2.1** *If  $(\Omega, \mathcal{F}, P)$  is a probability space and  $A_n \in \mathcal{F}$ ,  $n = 1, 2, \dots$ , are events on this space, then*

$$P(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n). \tag{2.4}$$

*Proof* Set  $B_1 = A_1$  and  $B_n = A_n \setminus (\cup_{i=1}^{n-1} A_i)$ ,  $n > 1$ . The sets  $\{B_n\}$  are disjoint events with the properties  $B_n \subseteq A_n$ , and  $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n$ . Hence,  $P(\cup_{n=1}^{\infty} A_n) = P(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n)$ .  $\blacktriangle$

**Definition 2.9** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $B \in \mathcal{F}$  an event such that  $P(B) > 0$ . The probability of  $A \in \mathcal{F}$  conditional on  $B$  is

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathcal{F}. \tag{2.5}$$

The definition is meaningful in the sense that  $P(\cdot | B)$  is a probability measure.

Suppose an experiment is performed in which  $B$  either occurs or does not occur, and that  $P(B), P(B^c) > 0$ . If  $B$  or  $B^c$  is observed, the conditional probability of  $A$

is  $P(A | B)$  or  $P(A | B^c)$ , respectively. Hence, the conditional probability of  $A$  is equal to  $P(A \cap B)/P(B)$  and  $P(A \cap B^c)/P(B^c)$  with probabilities  $P(B)$  and  $P(B^c)$ . This remark will be revisited later in this chapter (Example 2.56).

The following two formulas involving conditional probabilities are very useful for applications. Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $A_i \in \mathcal{F}$ ,  $i = 1, \dots, n$ , a partition of  $\Omega$ , that is,  $\Omega = \cup_{i=1}^n A_i$  and  $A_i \cap A_j = \emptyset$  for  $i \neq j$ . If  $P(A_i) > 0$ , and  $B \in \mathcal{F}$ , then

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B | A_i)P(A_i) \quad (\text{Law of total probability})$$

$$P(A_j | B) = \frac{P(A_j)P(B | A_j)}{P(B)} = \frac{P(A_j)P(B | A_j)}{\sum_{i=1}^n P(A_i)P(B | A_i)} \quad (\text{Bayes' formula}). \quad (2.6)$$

In the Bayesian framework, the probabilities  $P(A_j)$  and  $P(A_j | B)$  are referred to as the prior and the posterior probabilities of  $A_j$ .

*Example 2.4* Let  $B$  denote the event that a system performs satisfactorily, and let  $\{A_i, i = 1, 2\}$  be events partitioning  $\Omega$ . Suppose  $P(A_1) = 0.8$ ,  $P(A_2) = 0.2$ ,  $P(B | A_1) = 0.9$ , and  $P(B | A_2) = 0.7$ . The probability that the system performs satisfactorily is  $P(B) = (0.9)(0.8) + (0.7)(0.2) = 0.86$ .  $\diamond$

**Theorem 2.2** *The inequalities*

$$P\left(\cup_{i=1}^n A_i\right) \leq P_u = \sum_{i=1}^n P(A_i) - \sum_{i=2}^n \max_{j=1, \dots, i-1} P(A_j \cap A_i) \quad \text{and}$$

$$P\left(\cup_{i=1}^n A_i\right) \geq P_l = P(A_1) + \sum_{i=2}^n \max\left(0, P(A_i) - \sum_{j=1}^{i-1} P(A_j \cap A_i)\right), \quad (2.7)$$

hold for  $A_i \in \mathcal{F}$  arbitrary.

*Proof* Set  $B_i = A_i^c$ ,  $i = 1, \dots, n$ , and  $G = \cap_{i=1}^{n-1} B_i$ , and note that  $P\left(\cap_{i=1}^n B_i\right) = P(G \cap B_n) = P(G) - P(G \cap A_n)$ . Repeated application of this formula gives  $P\left(\cap_{i=1}^n B_i\right) = P(B_1) - \sum_{k=2}^n P(B_1 \cap \dots \cap B_{k-1} \cap A_k)$  so that

$$\begin{aligned} P\left(\cup_{i=1}^n A_i\right) &= 1 - P\left(\cap_{i=1}^n B_i\right) = P(A_1) + \sum_{k=2}^n P(B_1 \cap \dots \cap B_{k-1} \cap A_k) \\ &= P(A_1) + \sum_{k=2}^n P(B_1 \cap \dots \cap B_{k-1} | A_k)P(A_k). \end{aligned}$$

Since  $P(B_1 \cap \dots \cap B_{k-1} | A_k) \leq P(B_j | A_k) = 1 - P(A_j | A_k)$ ,  $j = 1, \dots, k-1$ , and  $P(B_1 \cap \dots \cap B_{k-1} | A_k) = 1 - P(A_1 \cup \dots \cup A_{k-1} | A_k) \geq 1 - \sum_{j=1}^{k-1} P(A_j | A_k)$  hold, we have the bounds in (2.7).  $\blacktriangle$

The bounds in (2.7) are relatively simple to calculate since they involve the probabilities of the events  $A_i$  and  $A_i \cap A_j$ , rather than probabilities of intersections of multiple events as in (2.3). Note that  $P(\cup_{i=1}^n A_i)$  can be interpreted as the probability of failure for a physical system with failure modes  $A_1, \dots, A_n$  occurring with probabilities  $P(A_i)$ ,  $i = 1, \dots, n$ .

### 2.2.4 Construction of Probability Spaces

In most applications a random experiment rather than a probability space is specified, so that we need to construct a probability space based on the available information. We present three methods for constructing probability spaces and illustrate them by examples.

#### 2.2.4.1 Countable Sample Space

Let  $\Omega = \{\omega_1, \omega_2, \dots\}$  be a countable sample space and let  $\{p_i \geq 0\}$  such that  $\sum_{i=1}^{\infty} p_i = 1$ . Take  $\mathcal{F}$  to be the collection of all subsets of  $\Omega$ , that is, the power set of  $\Omega$ . The set function  $P: \mathcal{F} \rightarrow [0, 1]$  defined by

$$P(A) = \sum_{\omega_i \in A} p_i, \quad A \in \mathcal{F}, \quad (2.8)$$

is a probability measure on  $(\Omega, \mathcal{F})$  since it is countably additive and has the properties  $P(\emptyset) = 0$  and  $P(\Omega) = 1$ .

*Example 2.5* Suppose service life has been recorded for  $n$  nominally identical devices. Let  $0 = \tau_0 < \tau_1 < \dots < \tau_{m-1}$ , let  $[0, \tau_1), [\tau_1, \tau_2), \dots, [\tau_{m-1}, \infty)$  be a partition of  $[0, \infty)$ , and denoted by  $n_i \leq n$  the number of devices with measured service life in the range  $[\tau_{i-1}, \tau_i)$ ,  $i = 1, \dots, m$ , with the notation  $\tau_m = \infty$ . The members of the probability space  $(\Omega, \mathcal{F}, P)$  describing this experiment are  $\Omega = \{\omega_1, \dots, \omega_m\}$ ,  $\mathcal{F}$  is the power set of  $\Omega$ , and  $P(\{\omega_i\}) \simeq n_i/n$ .  $\diamond$

#### 2.2.4.2 Product Probability Space

Let  $(\Omega_k, \mathcal{F}_k, P_k)$ ,  $k = 1, 2$ , be probability spaces. The elements of the product probability space  $(\Omega, \mathcal{F}, P)$  are

$$\begin{aligned} \Omega &= \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_k \in \Omega_k, k = 1, 2\} \\ \mathcal{F} &= \mathcal{F}_1 \times \mathcal{F}_2 = \sigma(\mathcal{R}), \quad \text{where } \mathcal{R} = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\} \\ P &= P_1 \times P_2, \quad \text{where } P(A_1 \times A_2) = P_1(A_1)P_2(A_2), \text{ for } A_1 \in \mathcal{F}_1 \text{ and } A_2 \in \mathcal{F}_2, \end{aligned} \quad (2.9)$$

and are called the product sample space,  $\sigma$ -field, and probability measure.

The product sample space  $\Omega$  contains the joint outcomes of, for example, two experiments. The  $\sigma$ -field  $\mathcal{F}$  can also be obtained from  $\mathcal{F} = \sigma(\mathcal{R}) = \sigma(\mathcal{G}_1, \mathcal{G}_2)$

where  $\mathcal{G}_1 = \{A_1 \times \Omega_2 : A_1 \in \mathcal{F}_1\}$  and  $\mathcal{G}_2 = \{\Omega_1 \times A_2 : A_2 \in \mathcal{F}_2\}$  since  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are  $\sigma$ -fields on  $\Omega$  that are included in  $\mathcal{F}$  and every member of  $\mathcal{R}$  is the intersection of sets from  $\mathcal{G}_1$  and  $\mathcal{G}_2$ . The construction of the product probability measure is less simple. It can be shown that there exists a unique probability  $P$  on  $(\Omega, \mathcal{F})$  that satisfies (2.9) ([4], Theorem 3.3.5).

The formulas in (2.9) can be generalized to define product probability spaces of  $n \geq 3$  probability spaces. The definition extends directly to the case in which  $n$  is infinity. If the probability spaces  $(\Omega_k, \mathcal{F}_k, P_k)$  are identical, the product sample space,  $\sigma$ -field, and probability are denoted by  $\Omega_1^n$ ,  $\mathcal{F}_1^n$ , and  $P_1^n$ .

*Example 2.6* Let  $(\Omega_k, \mathcal{F}_k, P_k)$ ,  $k = 1, 2$ , be probability spaces associated with the experiment of rolling two dice, so that  $\Omega_k = \{1, 2, 3, 4, 5, 6\}$ ,  $\mathcal{F}_k =$  all subsets of  $\Omega_k$ , and  $P_k(\{i\}) = 1/6$ ,  $i = 1, \dots, 6$ . The product sample space  $\Omega = \Omega_1 \times \Omega_2 = \{\omega = (i, j), i, j = 1, \dots, 6\}$  includes all outcomes of the experiment. The product  $\sigma$ -field  $\mathcal{F}$  consists of all subsets of  $\Omega$  since the members of  $\mathcal{R}$  are  $(i, j)$ ,  $\cup_{i \in I_1}(i, j)$ ,  $\cup_{j \in I_2}(i, j)$ ,  $\cup_{i \in I_1, j \in I_2}(i, j)$ , where  $I_1, I_2 \subseteq \{1, 2, 3, 4, 5, 6\}$ . The product probability measure is  $P(\{\omega\}) = 1/36$ ,  $\omega \in \Omega$ , since the outcomes of the experiment are equally likely.  $\diamond$

*Example 2.7* Consider a facility that fails under wind speeds exceeding a critical value  $v_{cr}$ . Let  $p = P(A)$ , where  $A = \{V > v_{cr}\}$  and  $V \geq 0$  denotes the maximum yearly wind speed at the facility site. The members of the probability space  $(\Omega, \mathcal{F}, P)$  describing this problem are  $\Omega = \{A, A^c\}$ ,  $\mathcal{F} =$  all parts of  $\Omega$ , and  $P(A) = p$ . To evaluate the probability that the facility performs satisfactorily during its design life of  $n$  years, we need to construct a new probability space corresponding to an “experiment” consisting of  $n$  independent repetitions of the yearly experiment. The construction resembles the experiment of tossing a loaded coin  $n$  times with sides  $\{1\}$  and  $\{0\}$  corresponding to events  $A$  and  $A^c$  ([11], p. 41).

The members of the product probability space corresponding to a  $n$  year time horizon are  $\Omega^n = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i = 0 \text{ or } 1\}$ ,  $\mathcal{F}^n =$  all subsets of  $\Omega^n$ , and  $P^n(B) = \sum_{\omega \in B} p^{n_\omega} q^{n-n_\omega}$ , where  $B \in \mathcal{F}^n$ ,  $q = 1 - p$ , and  $n_\omega = \sum_{i=1}^n \omega_i$  gives the numbers of 1's in  $\omega = (\omega_1, \dots, \omega_n)$ .  $\diamond$

### 2.2.4.3 Extension of Probability Measure

Let  $\Omega$  be a sample space associated with an experiment and let  $\mathcal{C}$  be a collection of subsets of  $\Omega$ . If  $\mathcal{C}$  is a field on  $\Omega$  and  $R$  is a real-valued, positive, and countably additive function defined on  $\mathcal{C}$  such that  $R(\Omega) = 1$ , then there exists a unique probability  $P$  on  $\mathcal{F} = \sigma(\mathcal{C})$  such that  $P(A) = R(A)$  for each  $A \in \mathcal{C}$ , that is, the restriction of  $P$  to  $\mathcal{C}$  is equal to  $R$  ([5], Theorem 14, p. 94). A field is a collection of sets satisfying the conditions of Definition 2.2 with  $I$  finite.

*Example 2.8* Let  $\Omega = \mathbb{R}$  and let  $\mathcal{C}$  consist of the empty set and the collection of all finite unions of intervals of the type  $(a, b]$  for  $a < b$ ,  $(-\infty, a]$ ,  $(a, \infty)$ , and  $(-\infty, \infty)$ . Let  $F: \mathbb{R} \rightarrow [0, 1]$  be a continuous increasing function such that

$\lim_{x \rightarrow -\infty} F(x) = 0$  and  $\lim_{x \rightarrow \infty} F(x) = 1$ , and define  $R: \mathcal{C} \rightarrow [0, 1]$  by  $R((a, b]) = F(b) - F(a)$ ,  $R((-\infty, a]) = F(a)$ ,  $R((a, \infty)) = 1 - F(a)$ , and  $R((-\infty, \infty)) = 1$ . For example,  $R((a, \infty))$  may represent the probability that the strength of a particular material exceeds  $a$ , so that it can be estimated from experiments. Since  $\mathcal{C}$  is a field, the set function  $R$  can be extended uniquely to a probability measure on  $(\mathbb{R}, \mathcal{B} = \sigma(\mathcal{C}))$  ([5], Proposition 9, p. 90, and Theorem 14, p. 94).  $\diamond$

## 2.3 Measurable Functions and Random Elements

Consider two measurable spaces  $(\Omega, \mathcal{F})$  and  $(\Psi, \mathcal{G})$  and a mapping  $h: \Omega \rightarrow \Psi$ .

**Definition 2.10** The mapping  $h: \Omega \rightarrow \Psi$  is said to be measurable from  $(\Omega, \mathcal{F})$  to  $(\Psi, \mathcal{G})$  or  $(\mathcal{F}, \mathcal{G})$ -measurable if  $h^{-1}(\mathcal{G}) \subseteq \mathcal{F}$ , that is,  $h^{-1}(B) = \{\omega \in \Omega : h(\omega) \in B\} \in \mathcal{F}$  for all  $B \in \mathcal{G}$ . This property of  $h$  is denoted by  $h \in \mathcal{F}/\mathcal{G}$  or just  $h \in \mathcal{F}$  if there is no confusion about  $\mathcal{G}$ .

*Example 2.9* Let  $A \in \mathcal{F}$ ,  $(\Psi, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$ , and  $1_A$  the indicator function of  $A$ , that is,  $1_A(\omega) = 1$  for  $\omega \in A$  and  $1_A(\omega) = 0$  for  $\omega \notin A$ . The function  $1_A: \Omega \rightarrow \mathbb{R}$  is  $\mathcal{F}/\mathcal{B}$ -measurable since  $1_A^{-1}(\{0\}) = A^c$  and  $1_A^{-1}(\{1\}) = A$ , so that  $1_A^{-1}(\mathcal{B}) = \{\emptyset, \Omega, A, A^c\} \subseteq \mathcal{F}$ .  $\diamond$

*Example 2.10* Let  $\{A_i\}$ ,  $i = 1, \dots, n$ , be events partitioning  $\Omega$  and  $\{c_i\}$ ,  $i = 1, \dots, n$ , real constants. The function  $h = \sum_{i=1}^n c_i 1_{A_i}$  is  $(\mathcal{F}, \mathcal{G})$ -measurable, where  $\Psi = \{c_1, \dots, c_n\}$  and  $\mathcal{G}$  is the power set of  $\Psi$ . The image of the members of  $\Psi$  in  $\Omega$  can be obtained simply, for example,  $h^{-1}(\{c_i, c_j\}) = A_i \cup A_j \in \mathcal{F}$ .  $\diamond$

**Theorem 2.3** If  $h: \mathbb{R}^d \rightarrow \mathbb{R}^q$  is continuous, it is also  $(\mathcal{B}^d, \mathcal{B}^q)$ -measurable.

*Proof* Function  $h$  is measurable if and only if  $h^{-1}(\mathcal{T}^q) \subseteq \mathcal{B}^d$  ([11], Proposition 3.2.1), where  $\mathcal{T}^r$  denotes the topology generated by the open balls in  $\mathbb{R}^r$ . Since  $h$  is continuous, we have  $h^{-1}(\mathcal{T}^q) \subseteq \mathcal{T}^d \subseteq \mathcal{B}^d$  (Sect. B.1.1), where the latter inequality holds since  $\mathcal{B}^d$  is generated by  $\mathcal{T}^d$ .  $\blacktriangle$

**Definition 2.11** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $(S, \mathcal{S})$  a measurable space, where  $S$  is a metric space and  $\mathcal{S} = \sigma(\mathcal{T})$  denotes the  $\sigma$ -field generated by the topology  $\mathcal{T}$  induced on  $S$  by its metric (Example 2.3 and Sect. B.1). The mapping  $X: \Omega \rightarrow S$  is a random element if it is  $(\mathcal{F}, \mathcal{S})$ -measurable.

The element  $X$  in the above definition is a real-valued random variable, complex-valued random variable, random vector or an  $\mathbb{R}^d$ -valued random variable,  $\mathbb{C}^d$ -valued random variable, real-valued stochastic process with continuous samples defined on  $[0, 1]$ , or a real-valued random field with continuous samples defined on a subset  $D$  of  $\mathbb{R}^{d'}$  if  $S = \mathbb{R}$ ,  $S = \mathbb{C}$ ,  $S = \mathbb{R}^d$ ,  $S = \mathbb{C}^d$ ,  $S = C[0, 1]$ , or  $S = C(D)$ , respectively. The image  $X(\omega)$ ,  $\omega \in \Omega$ , of  $X$  is a number, vector in  $\mathbb{R}^d$ , vector in  $\mathbb{C}^d$ , real-valued continuous function defined on  $[0, 1]$ , or a real-valued continuous function defined on  $D$  if  $S = \mathbb{R}$ ,  $S = \mathbb{R}^d$ ,  $S = \mathbb{C}^d$ ,  $S = C[0, 1]$ , or  $S = C(D)$ , respectively. For

example,  $X(\omega) = (X_1(\omega), \dots, X_d(\omega)) \in \mathbb{R}^d$  is a  $d$ -dimensional vector for  $S = \mathbb{R}^d$  and  $X(\omega)(t)$ ,  $t \in [0, 1]$ , is a sample of a real-valued stochastic process defined on  $[0, 1]$  for  $S = C[0, 1]$ . It is common to denote  $X(\omega)(t)$  by  $X(t, \omega)$ , as will be seen in a subsequent section. For a fixed time  $t \in [0, 1]$ ,  $X(t, \omega)$  is a random variable.

**Definition 2.12** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $(S, \mathcal{S})$  a measurable space, and  $X : \Omega \rightarrow S$  an  $(\mathcal{F}, \mathcal{S})$ -measurable mapping. The probability measure induced by  $X$  or the distribution of  $X$  is the probability measure

$$Q(B) = P(X^{-1}(B)) = P \circ X^{-1}(B), \quad B \in \mathcal{S}, \quad (2.10)$$

on the measurable space  $(S, \mathcal{S})$ .

*Example 2.11* If  $S = \mathbb{R}$  and  $B = (-\infty, x]$ ,  $x \in \mathbb{R}$ , then  $Q(B) = P(\{\omega : X(\omega) \leq x\}) = P(X \leq x)$  is called the probability distribution function or just the distribution of random variable  $X$ , and is typically denoted by  $F(x) = P(X \leq x)$ . If  $S = \mathbb{R}^d$  and  $B = \times_{i=1}^d (-\infty, x_i]$ ,  $x_i \in \mathbb{R}$ , then  $Q(B) = P(\{\omega : X_i(\omega) \leq x_i, i = 1, \dots, d\}) = P(\cap_{i=1}^d \{X_i \leq x_i\})$  is called the joint distribution of random vector  $X$ . If  $S = C[0, 1]$ ,  $\{t_1, \dots, t_n\} \subset [0, 1]$ , and  $B = \times_{k=1}^n (-\infty, x_k]$ ,  $x_k \in \mathbb{R}$ , then  $Q(B) = P(\{\omega : X(t_k, \omega) \leq x_k, k = 1, \dots, n\}) = P(\cap_{k=1}^n \{X(t_k) \leq x_k\})$  is called the finite dimensional distribution of order  $n$  of  $X$ . The finite dimensional distributions of a stochastic process  $X$  are the joint distributions of random vectors  $(X(t_1), \dots, X(t_n))$  consisting of values of the process at times  $t_1, \dots, t_n$ .  $\diamond$

*Example 2.12* Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $A \in \mathcal{F}$ ,  $S = \{0, 1\}$ , and  $1_A : \Omega \rightarrow S$  an indicator function. The function  $1_A$  is  $(\mathcal{F}, \mathcal{S})$ -measurable, where  $\mathcal{S} = \{\emptyset, S, \{0\}, \{1\}\}$ . The probability measure  $Q$  induced by  $1_A$  on  $(S, \mathcal{S})$  is  $Q(\{0\}) = P(1_A^{-1}(\{0\})) = P(A^c)$  and  $Q(\{1\}) = P(1_A^{-1}(\{1\})) = P(A)$ . The distribution of random variable  $1_A$  is  $F(x) = P(A^c)1(x \geq 0) + P(A)1(x \geq 1)$ .  $\diamond$

*Example 2.13* Let  $(\Omega, \mathcal{F}, P)$  be a probability space and let  $X : \Omega \rightarrow \mathbb{R}$  be  $(\mathcal{F}, \mathcal{B})$ -measurable. Suppose the probability measure  $Q$  induced by  $P$  on  $(\mathbb{R}, \mathcal{B})$  has the expression  $Q((-\infty, x]) = P(X^{-1}((-\infty, x])) = \Phi((x - \mu)/\sigma)$ , where  $x, \mu$ , and  $\sigma > 0$  are reals,  $\Phi(u) = \int_{-\infty}^u \phi(\xi) d\xi$ , and  $\phi(\xi) = \exp(-\xi^2/2)/\sqrt{2\pi}$ . Then  $X$  is said to be Gaussian random variable with mean  $\mu$ , standard deviation  $\sigma$ , and variance  $\sigma^2$ , a property denoted by  $X \sim N(\mu, \sigma^2)$ .  $\diamond$

**Theorem 2.4** Let  $(\Omega, \mathcal{F}, P)$  be a probability space. A function  $X : \Omega \rightarrow \mathbb{R}^d$  is an  $\mathbb{R}^d$ -valued random variable on  $(\Omega, \mathcal{F}, P)$  if its coordinates are real-valued random variables on this probability space.

*Proof* If  $X$  is a random vector, its coordinates  $X_i = \pi_i \circ X$ ,  $i = 1, \dots, d$ , are random variables because the projection map  $\pi_i(x) = x_i$  is continuous.

Suppose now that  $X_i$  are random variables. Let  $\mathcal{R}$  be the collection of open rectangles in  $\mathbb{R}^d$  with members  $R = I_1 \times \dots \times I_d$ , where  $I_i$  are open intervals in  $\mathbb{R}$ . The  $\sigma$ -field  $\mathcal{B}^d$  is generated by these rectangles, that is,  $\mathcal{B}^d = \sigma(\mathcal{R})$  ([11], Proposition 3.2.4), so that it is sufficient to show that  $X^{-1}(\mathcal{R})$  is in  $\mathcal{F}$ . We have

$X^{-1}(R) = \cap_{i=1}^d X_i^{-1}(I_i) \in \mathcal{F}$  since  $X_i$  are random variables. Hence,  $X$  is a random vector if and only if  $\{\omega \in \Omega : X_i(\omega) \leq x_i\} \in \mathcal{F}$  for all  $x_i \in \mathbb{R}$ ,  $i = 1, \dots, d$ .  $\blacktriangle$

*Example 2.14* Consider a series  $X = (X_1, X_2, \dots)$ , where  $X_i$  are measurable functions from  $(\Omega, \mathcal{F})$  to  $(\Psi, \mathcal{G})$ . Let  $\mathcal{K}$  be the collection of all subsets of  $\mathbb{Z}^+ = \{1, 2, \dots\}$ . The function  $(m, \omega) \mapsto X_m(\omega)$  depending on the arguments  $m$  and  $\omega$  is measurable from  $(\mathbb{Z}^+ \times \Omega, \mathcal{K} \times \mathcal{F})$  to  $(\Psi, \mathcal{G})$ . Generally, this property does not hold if the discrete index  $m$  in this example is allowed to take values in an uncountable set.  $\diamond$

*Proof* Let  $A = \{(m, \omega) : X_m(\omega) \in B\}$  be the inverse image of the function  $(m, \omega) \mapsto X_m(\omega)$  in  $\mathbb{Z}^+ \times \Omega$  corresponding to an arbitrary member  $B$  of  $\mathcal{G}$ . Because  $X_m$  is measurable, the set  $A$  can be expressed as the countable union  $\cup_m \{\omega : X_m(\omega) \in B\}$  of sets  $\{\omega : X_m(\omega) \in B\}$  that are in  $\mathcal{F}$  for each  $m \geq 1$ . Hence,  $A$  is in  $\mathcal{K} \times \mathcal{F}$ . Note also that the function  $m \mapsto X_m(\omega)$  is  $\mathcal{K}$ -measurable for a fixed  $\omega \in \Omega$  since  $\{m : X_m(\omega) \in B\}$  is a subset of  $\mathbb{Z}^+$  so that it is in  $\mathcal{K}$ .  $\blacktriangle$

**Definition 2.13** Let  $(\Omega, \mathcal{F})$  and  $(S, \mathcal{S})$  be measurable spaces and let  $X : \Omega \rightarrow S$  be a random element, that is, an  $(\mathcal{F}, \mathcal{S})$ -measurable mapping. The  $\sigma$ -field generated by  $X$ , denoted by  $\sigma(X)$  or  $\mathcal{F}^X$ , is  $\sigma(X) = X^{-1}(\mathcal{S}) = \{X^{-1}(B) : \forall B \in \mathcal{S}\}$ .

*Example 2.15* Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $1_A : \Omega \rightarrow \mathbb{R}$ ,  $A \in \mathcal{F}$ , an indicator function. The  $\sigma$ -field generated by  $1_A$  is  $\sigma(1_A) = \{\emptyset, \Omega, A, A^c\}$ . There is no smaller field with respect to which  $1_A$  is measurable.  $\diamond$

It is common in applications to be interested in functions of random variables. For example, we may have to find the distribution of the deformation or any other response  $Y$  of a physical system with random properties  $X_1$  that is subjected to a random action  $X_2$ . Suppose  $X_1$  and  $X_2$  are random variables on a probability space  $(\Omega, \mathcal{F}, P)$ . To achieve this objective, we first need to show that  $Y$ , which is a function of  $(X_1, X_2)$ , is a random variable on  $(\Omega, \mathcal{F}, P)$ . The following theorem shows that  $Y$  has this property if the mapping  $(X_1, X_2) \mapsto Y$  is measurable.

**Theorem 2.5** Let  $h : (\Omega, \mathcal{F}) \rightarrow (\Psi, \mathcal{G})$  and  $g : (\Psi, \mathcal{G}) \rightarrow (\Phi, \mathcal{H})$  be measurable functions, where  $(\Omega, \mathcal{F})$ ,  $(\Psi, \mathcal{G})$ ,  $(\Phi, \mathcal{H})$  are measurable spaces. Then the function  $g \circ h : (\Omega, \mathcal{F}) \rightarrow (\Phi, \mathcal{H})$  is measurable.

*Proof* We have  $g^{-1}(\mathcal{H}) \subset \mathcal{G}$  and  $h^{-1}(g^{-1}(\mathcal{H})) \subset h^{-1}(\mathcal{G}) \subset \mathcal{F}$  since  $g$  and  $h$  are measurable functions.  $\blacktriangle$

## 2.4 Independence

We define independent  $\sigma$ -fields and show that this concept applies directly to characterize independent events and random elements.

**Definition 2.14** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\mathcal{F}_i$ ,  $i \in I$ , a collection of sub- $\sigma$ -fields of  $\mathcal{F}$ . If  $I$  is finite, the  $\sigma$ -fields  $\mathcal{F}_i$ ,  $i \in I$ , are independent if

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i), \quad \forall A_i \in \mathcal{F}_i. \quad (2.11)$$

If  $I$  is infinite and (2.11) holds for all finite subsets of  $I$ , then the  $\sigma$ -fields  $\mathcal{F}_i$ ,  $i \in I$ , are said to be independent.

The definition implies that (2.11) must hold for any subset of  $\{A_i \in \mathcal{F}_i, i \in I\}$  since some  $A_i$  can be selected to coincide with  $\Omega$ . If the  $\sigma$ -fields  $\mathcal{F}_i$ ,  $i \in I$ , are on different sample spaces, the above independence condition needs to be applied on the corresponding product probability space.

**Definition 2.15** A collection of events  $A_i \in \mathcal{F}$ ,  $i \in I$ , is said to be independent if the  $\sigma$ -fields  $\sigma(A_i) = \{\emptyset, \Omega, A_i, A_i^c\}$  are independent.

*Example 2.16* Two events  $A$  and  $B$  on a probability space  $(\Omega, \mathcal{F}, P)$  are independent if  $P(A \cap B) = P(A)P(B)$ . This is the classical definition for the independence between two events.  $\diamond$

*Proof* By definition,  $A$  and  $B$  are independent if the fields  $\sigma(A)$  and  $\sigma(B)$  are independent, which implies  $P(A' \cap B') = P(A')P(B')$  for all  $A' \in \sigma(A)$  and  $B' \in \sigma(B)$ . The stated condition of independence follows by considering all possible pairs of events  $(A', B')$ . The converse also holds, that is,  $P(A \cap B) = P(A)P(B)$  implies the independence of the fields  $\sigma(A)$  and  $\sigma(B)$ . For example,  $P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(A^c)P(B)$ .

Note also that, the classical definition of independence between two events follows from that of the conditional probability. If  $A$  and  $B$  are independent and  $P(B) > 0$ , the conditional probability  $P(A | B) = P(A \cap B)/P(B)$  is unaffected by the occurrence of  $B$  so that  $P(A | B) = P(A)$  implying  $P(A \cap B) = P(A)P(B)$ .  $\blacktriangle$

*Example 2.17* Consider two events  $A$  and  $B$  as in Example 2.16. If  $A \cap B = \emptyset$  and  $P(A), P(B) > 0$ , then  $A$  and  $B$  are not independent since  $P(A \cap B) = P(\emptyset) = 0$  and  $P(A), P(B) > 0$ .  $\diamond$

**Definition 2.16** The events  $A_i$ ,  $i = 1, \dots, n$ , on a probability space  $(\Omega, \mathcal{F}, P)$  are independent if

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = \prod_{k=1}^m P(A_{i_k}) \quad (2.12)$$

holds for any subset  $\{i_1, \dots, i_m\}$  of  $\{1, \dots, n\}$ .

*Example 2.18* The requirement in (2.12) is essential for three or more events. It is not sufficient to satisfy (2.12) for the entire collection of events. For example, set  $\Omega = \{1, 2, 3, 4\}$ ,  $\mathcal{F} =$  all parts of  $\Omega$ , and  $P(\{1\}) = \sqrt{2}/2 - 1/4$ ,  $P(\{2\}) = 1/4$ ,  $P(\{3\}) = 3/4 - \sqrt{2}/2$ , and  $P(\{4\}) = 1/4$ . Let  $A_1 = \{1, 3\}$ ,  $A_2 = \{2, 3\}$ , and  $A_3 = \{3, 4\}$  be events in  $(\Omega, \mathcal{F})$ . The probability of  $A_1 \cap A_2 \cap A_3 = \{3\}$  is  $P(\{3\}) = 3/4 - \sqrt{2}/2$  and is equal to  $P(A_1)P(A_2)P(A_3)$ . However,  $P(A_1 \cap A_2) \neq P(A_1)P(A_2)$  ([9], p. 2).  $\diamond$



**Definition 2.17** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and let  $\mathcal{C}_i \subset \mathcal{F}$ ,  $i \in I$ , be families of subsets of  $\mathcal{F}$ . If  $I = \{1, \dots, n\}$  is finite,  $\mathcal{C}_i$  are said to be independent if any  $A_1 \in \mathcal{C}_1, \dots, A_n \in \mathcal{C}_n$  are independent events. If  $I$  is not finite,  $\mathcal{C}_i$  are independent if  $\mathcal{C}_i$ ,  $i \in J$ , are independent families for each finite  $J \subseteq I$ .

This definition and the following criterion can be used to show that two or more  $\sigma$ -fields are independent. The criterion uses classes of events forming a  $\pi$ -system. A collection  $\mathcal{C}$  of subsets of  $\Omega$  is said to be a  $\pi$ -system if it is closed to finite intersection, that is,  $A, B \in \mathcal{C}$  implies  $A \cap B \in \mathcal{C}$ . If  $\mathcal{C}_i$  is a nonempty class of events in  $\mathcal{F}$  for each  $i = 1, \dots, n$ , such that (1)  $\mathcal{C}_i$  is a  $\pi$ -system and (2)  $\mathcal{C}_i$ ,  $i = 1, \dots, n$ , are independent, then the  $\sigma$ -fields  $\sigma(\mathcal{C}_i)$ ,  $i = 1, \dots, n$ , are independent ([11], Theorem 4.1.1).

**Definition 2.18** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $S$  a metric space,  $\mathcal{S}$  the  $\sigma$ -field induced on  $S$  by its metric, and  $X_i : \Omega \rightarrow S$ ,  $i \in I$ , a collection of random elements, that is, a collection of  $(\mathcal{F}, \mathcal{S})$ -measurable functions, where  $I$  may or may not be finite. The random elements  $X_i$  are independent if the  $\sigma$ -fields  $\sigma(X_i) = \sigma(X_i^{-1}(\mathcal{S}))$ ,  $i \in I$  generated by these elements are independent.

*Example 2.19* Let  $X$  and  $Y$  be real-valued random variables defined on  $(\Omega, \mathcal{F}, P)$ . The independence of  $\sigma(X)$  and  $\sigma(Y)$  implies  $P(X^{-1}((-\infty, x]) \cap Y^{-1}((-\infty, y])) = P(X \leq x)P(Y \leq y) = F_X(x)F_Y(y)$ ,  $x, y \in \mathbb{R}$ , where  $F_X$  and  $F_Y$  denote the distributions of  $X$  and  $Y$ . The converse also holds, that is,  $P(X \leq x, Y \leq y) = F_X(x)F_Y(y)$  implies the independence of  $\sigma(X)$  and  $\sigma(Y)$ . For example,  $P(\{a_1 < X \leq a_2\} \cap \{b_1 < Y \leq b_2\}) = P(X \leq a_2, Y \leq b_2) - P(X \leq a_1, Y \leq b_2) - P(X \leq a_2, Y \leq b_1) + P(X \leq a_1, Y \leq b_1)$  by properties of the probability measure, or  $P(\{a_1 < X \leq a_2\} \cap \{b_1 < Y \leq b_2\}) = F_X(a_2)F_Y(b_2) - F_X(a_1)F_Y(b_2) - F_X(a_2)F_Y(b_1) + F_X(a_1)F_Y(b_1)$ , so that  $P(\{a_1 < X \leq a_2\} \cap \{b_1 < Y \leq b_2\}) = (F_X(a_2) - F_X(a_1))(F_Y(b_2) - F_Y(b_1)) = P(a_1 < X \leq a_2)P(b_1 < Y \leq b_2)$ . The latter relationship implies the independence of the  $\sigma$ -fields  $\sigma(X)$  and  $\sigma(Y)$  since the intervals  $(a_1, a_2]$  and  $(b_1, b_2]$  are arbitrary.  $\diamond$

*Example 2.20* Let  $X_k$ ,  $k = 1, 2, \dots$ , be independent, real-valued random variables on a probability space  $(\Omega, \mathcal{F}, P)$  and  $\varphi_k : \mathbb{R} \rightarrow \mathbb{R}$  Borel measurable functions. The random variables  $\varphi_k \circ X_k$ ,  $k = 1, 2, \dots$ , are independent.

*Proof* We have  $\varphi_k^{-1}(\mathcal{B}) \subseteq \mathcal{B}$  and  $X_k^{-1}(\varphi_k^{-1}(\mathcal{B})) \subseteq \mathcal{F}$  because  $\varphi_k$  is a Borel measurable function and  $X_k$  is a random variable. Since  $X_k^{-1}(\varphi_k^{-1}(\mathcal{B})) \subseteq X_k^{-1}(\mathcal{B})$  and the  $\sigma$ -fields  $X_k^{-1}(\mathcal{B})$ ,  $k = 1, 2, \dots$ , are independent by assumption, the random variables  $\varphi_k \circ X_k$  are independent.  $\diamond$

*Example 2.21* Let  $X(t)$  and  $Y(t)$ ,  $t \in [0, 1]$ , be simple real-valued processes with continuous samples  $(x_1(t), \dots, x_m(t))$  and  $(y_1(t), \dots, y_n(t))$  occurring with probabilities  $(p_1, \dots, p_m)$  and  $(q_1, \dots, q_n)$ , respectively. It is assumed that both processes are defined on the same probability space  $(\Omega, \mathcal{F}, P)$ , so that they are measurable functions from  $(\Omega, \mathcal{F})$  to  $(C[0, 1], \mathcal{C})$ , and  $A_i = \{\omega \in \Omega : X(\omega) = x_i\}$ ,  $i = 1, \dots, m$ , and  $B_j = \{\omega \in \Omega : Y(\omega) = y_j\}$ ,  $j = 1, \dots, n$ , are measurable partitions of  $\Omega$ . The

processes  $X$  and  $Y$  are independent if the  $\sigma$ -fields generated by  $\{A_i, i = 1, \dots, m\}$  and  $\{B_j, j = 1, \dots, n\}$  are independent.  $\diamond$

## 2.5 Sequence of Events

Let  $\{A_n, n = 1, 2, \dots\}$  be a sequence of events in a probability space  $(\Omega, \mathcal{F}, P)$ . Properties of probability measures that involve increasing/decreasing, convergent, and arbitrary sequences of events are discussed.

**Definition 2.19** The sequence  $\{A_n, n = 1, 2, \dots\}$  is said to be increasing if  $A_n \subseteq A_{n+1}$  for all  $n$ . If  $A_n \supseteq A_{n+1}$  for all  $n$ , the sequence is decreasing. The sequence is convergent if  $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$  and  $\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$  coincide, and we use the notation  $\lim_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n$  for the limit of  $\{A_n\}$ . Note that  $\limsup_{n \rightarrow \infty} A_n$ ,  $\liminf_{n \rightarrow \infty} A_n$ , and  $\lim_{n \rightarrow \infty} A_n$  are events in  $(\Omega, \mathcal{F}, P)$ .

**Theorem 2.6** (Continuity of probability measure) *Let  $\{A_n, n = 1, 2, \dots\}$  be an increasing or decreasing sequence of events. The numerical sequence  $\{P(A_n), n = 1, 2, \dots\}$  is increasing or decreasing and converges to  $P(A)$ , where  $A = \lim_{n \rightarrow \infty} A_n$ .*

*Proof* Suppose  $\{A_n\}$  is increasing, so that it converges to  $A = \lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$ . Set  $B_1 = A_1$  and  $B_n = A_n \setminus A_{n-1}$ ,  $n = 2, 3, \dots$ , so that  $A_n = \bigcup_{k=1}^n B_k$ ,  $A = \bigcup_{n=1}^{\infty} B_n$ ,  $P(A) = P(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(B_k) = \lim_{n \rightarrow \infty} P(\bigcup_{k=1}^n B_k) = \lim_{n \rightarrow \infty} P(A_n)$  since  $\{B_n\}$  are disjoint events. Similar arguments hold for decreasing sequences.  $\blacktriangle$

A direct consequence of Theorem 2.6 is that, for a sequence  $\{A_n, n = 1, 2, \dots\}$  of convergent events, probability and limit can be interchanged, that is,

$$\lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n) = P(A), \quad (2.13)$$

where  $A = \limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n$  (Exercise 2.9 and [11], Sect. 2.1).

Let  $\{A_n, n = 1, 2, \dots\}$  be a sequence of events in a probability space  $(\Omega, \mathcal{F}, P)$ , and let  $A = \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$  be an event in this space. It can be seen that  $\omega \in A$  if and only if  $\omega \in A_n$  for infinitely many indices  $n$ . We use the notation

$$\{A_n \text{ i.o.}\} = \{A_n \text{ infinitely often}\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \limsup_{n \rightarrow \infty} A_n \quad (2.14)$$

to indicate this property.

**Theorem 2.7** (Borel–Cantelli lemma) *If  $\{A_n, n = 1, 2, \dots\}$  is a sequence of events such that  $\sum_{n=1}^{\infty} P(A_n) < \infty$ , then  $P(A_n \text{ i.o.}) = 0$ .*

*If  $\{A_n, n = 1, 2, \dots\}$  is a sequence of independent events, then  $P(A_n \text{ i.o.}) = 0$  and  $P(A_n \text{ i.o.}) = 1$  if and only if  $\sum_{n=1}^{\infty} P(A_n) < \infty$  and  $\sum_{n=1}^{\infty} P(A_n) = \infty$ , respectively.*

*Proof* We have  $P(A_n \text{ i.o.}) = \lim_{n \rightarrow \infty} P(\cup_{k=n}^{\infty} A_k) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) = 0$ , where the first and the last equalities hold since  $\{\cup_{k=n}^{\infty} A_k\}$  is an increasing sequence of events and  $\sum_{n=1}^{\infty} P(A_n)$  is convergent by assumption.

The proof of the second part of the Borel–Cantelli lemma can be found in [11] (Proposition 4.5.2). ▲

*Example 2.22* Let  $\{X_n, n = 1, 2, \dots\}$  be a sequence of Bernoulli random variables taking the values 1 and 0 with probabilities  $P(X_n = 1) = p_n = 1 - P(X_n = 0)$ ,  $n \geq 1$ . If  $\sum_{n=1}^{\infty} p_n < \infty$ , then  $P(\{X_n = 1\} \text{ i.o.}) = 0$  by the Borel–Cantelli lemma, so that  $X_n$  is equal to 1, a finite number of times. Other illustrations of the use of the Borel–Cantelli lemma can be found in [11] (Sect. 4.5). ◇

*Example 2.23* Let  $p_0 = P(V \leq v_{\text{cr}})$  denote the probability that yearly wind speed maximum does not exceed a critical value  $v_{\text{cr}}$  at a site. The event  $\{V \leq v_{\text{cr}}\}$  occurs infinitely often with probability one since  $\sum_{n=1}^{\infty} p_0 = \infty$  for  $p_0 > 0$  and the events  $\{V \leq v_{\text{cr}}\}$  in distinct years are independent. The statement follows from the second part of the Borel–Cantelli lemma. ◇

## 2.6 Expectation

The expectation operator is first defined for simple random variables. The definition is then extended to positive and arbitrary random variables, random vectors, and random functions. The expectation operator is used to define moments of random elements. Fubini’s theorem and applications of this theorem conclude the section.

**Definition 2.20** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $\{A_i \in \mathcal{F}, i \in I\}$  a partition of  $\Omega$ ,  $I$  a finite index set, and  $a_i \in \mathbb{R}^d$  such that  $\|a_i\| < \infty$ . Then

$$X = \sum_{i \in I} a_i 1_{A_i}, \quad a_i \in \mathbb{R}^d, \quad (2.15)$$

is called a finite, simple  $\mathbb{R}^d$ -valued random variable. The collection of simple random variable is a vector space (Exercise 2.11).

**Definition 2.21** The expectation of  $X$  in (2.15), denoted by  $E[X]$ ,  $\int_{\Omega} X dP$ , or  $\int_{\Omega} X(\omega) dP(\omega)$ , is

$$E[X] = \int_{\Omega} X dP = \int_{\Omega} X(\omega) dP(\omega) = \sum_{i \in I} a_i P(A_i). \quad (2.16)$$

Consider a mapping  $g: \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$  such that  $\|g(a_i)\| < \infty$ ,  $i \in I$ . Then,  $g(X)$  is a simple  $\mathbb{R}^{d'}$ -valued random variable with expectation

$$E[g(X)] = \sum_{i \in I} g(a_i) P(A_i). \quad (2.17)$$

The definition is meaningful since  $\|g(a_i)\| < \infty$ ,  $i \in I$ .

Following are the properties of the expectation for simple random variables. Except for Jensen's inequality (Exercise 2.12), all the other properties result directly from (2.16) and (2.17). The random variables in the first three properties given by (2.18) are real-valued.

If  $a_i \geq 0$  in (2.15), then  $E[X] \geq 0$ ,

If  $X \leq Y$  a.s., then  $E[X] \leq E[Y]$ ,

If  $g: \mathbb{R} \rightarrow \mathbb{R}$  is convex, then  $g(E[X]) \leq E[g(X)]$  (Jensen's inequality),

$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$ ,  $\alpha, \beta \in \mathbb{R}$ , and

$$\int_{\cup_i B_i} X dP = \sum_i \int_{B_i} X dP, \quad \text{where } \{B_i \in \mathcal{F}\} \text{ is a measurable partition of } \Omega. \quad (2.18)$$

The second and the fourth properties show that  $E[\cdot]$  is a monotone and a linear operator. Jensen's inequality implies  $|E[X]| \leq E[|X|]$  since the absolute value is a convex function. The inequality also follows from the definition of the expectation, which gives  $|E[X]| = |\sum_{i \in I} a_i P(A_i)| \leq \sum_{i \in I} |a_i| P(A_i) = E[|X|]$ .

We now extend the definition of expectation in (2.16) to positive real-valued random variables and, subsequently, arbitrary random variables.

**Definition 2.22** Let  $X$  be a real-valued, positive random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ , that is,  $P(X \geq 0) = 1$ . If  $P(X = \infty) > 0$ , set  $E[X] = \infty$ . Otherwise,

$$E[X] = \lim_{n \rightarrow \infty} E[X_n], \quad (2.19)$$

where  $X_n$  is an approximating sequence of finite-valued simple random variables such that  $X_n \uparrow X$  and  $E[X_n]$  is finite for each  $n$ .

The definition is meaningful since (1) there exists an increasing sequence of simple random variables  $X_n \geq 0$ ,  $n = 1, 2, \dots$ , referred to as an approximating sequence of  $X$ , such that  $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$  for almost all  $\omega$ 's ([11], Theorem 5.1.1) and (2) the expectation in (2.19) is well-defined since the value of  $E[X]$  does not change if  $\{X_n\}$  is replaced with another approximating sequence  $Y_m \uparrow X$  ([11], Proposition 5.2.1).

**Definition 2.23** Let  $X$  be a real-valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ , and set  $X^+ = X \vee 0 = \max(X, 0)$  and  $X^- = (-X) \vee 0 = \max(-X, 0)$ . The expectation of  $X$  is

$$E[X] = \begin{cases} E[X] = E[X^+] - E[X^-], & \text{if } E[X^+] < \infty \text{ and/or } E[X^-] < \infty \\ \text{Does not exist,} & \text{if } E[X^+] = E[X^-] = \infty. \end{cases} \quad (2.20)$$

If both expectations  $E[X^+]$  and  $E[X^-]$  are finite, then  $E[X]$  exists, is finite, and has the expression  $E[X] = E[X^+] - E[X^-]$ . If only one of the expectations  $E[X^+]$  and

$E[X^-]$  is finite, then  $E[X]$  exists but is unbounded. As for simple random variables, we have

$$E[X1_B] = \int_{\Omega} 1_B(\omega)X(\omega) dP(\omega) = \int_B X dP, \quad B \in \mathcal{F}, \quad (2.21)$$

and, if  $E[X1_B]$  is finite, we say that  $X$  is integrable with respect to  $P$  over  $B$  or just  $P$ -integrable over  $B$ . The definitions in (2.20) and (2.21) extend directly to random vectors and complex-valued random variables by applying them to the coordinates and real/imaginary parts of these variables, respectively.

**Example 2.24** Let  $X(\omega) = i^2 + j^2$  be a random variable defined on  $(\Omega, \mathcal{F}, P)$ , where  $\Omega = \{\omega = (i, j) : i, j = 1, 2, \dots, 6\}$  is the sample space for the experiment of rolling two dice,  $\mathcal{F}$  consists of all subsets of  $\Omega$ , and  $P(\{\omega\}) = 1/36$ . Then  $X$  is a positive, simple random variable and  $E[X1_B] = \sum_{\omega \in B} (i^2 + j^2)(1/36) = [(2)(40) + (2)(34) + 32]/36 = 5$  for  $B = \{\omega = (i, j) : i + j = 8\}$ .  $\diamond$

The second and the third properties in (2.18) are also valid for the expectation in (2.20) (Exercises 2.12–2.14 and [4], Sect. 3.2). Following are the properties specific to the expectation in (2.20).

**Theorem 2.8** *Let  $X$  be a real-valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Then ([4], Sect. 3.2)*

$$E[X1_A] = \int_A X dP, \quad A \in \mathcal{F}, \quad \text{is finite if and only if } \int_A |X| dP < \infty,$$

$$\text{If } X \geq 0 \text{ a.s., then } \int_A X dP \geq 0, \text{ and}$$

*If  $X$  is  $P$ -integrable, then  $X$  is finite a.s., that is,*

$$N = \{\omega : X(\omega) = \pm\infty\} \in \mathcal{F} \text{ and } P(N) = 0. \quad (2.22)$$

**Definition 2.24** Let  $X$  be a real-valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$  and let  $q \geq 1$  be an integer. If  $\mu(q) = E[X^q]$  exists and is finite, it is called the moment of order  $q$  of  $X$ . The mean of  $X$  is  $\mu = \mu(1)$ . If  $E[(X - \mu)^q]$  exists and is finite, it is called the central moment of order  $q$  of  $X$ . The central moment  $\sigma^2 = E[(X - \mu)^2]$  is the variance of  $X$ . The square root  $\sigma$  of the variance  $\sigma^2$  is the standard deviation of  $X$ . The scaled versions of the third and fourth central moments  $\gamma_3 = E[(X - \mu)^3]/\sigma^3$  and  $\gamma_4 = E[(X - \mu)^4]/\sigma^4$  are the skewness and kurtosis coefficients of  $X$ . The ratio  $v = \sigma/\mu$  defined for  $\mu \neq 0$  is called coefficient of variation.

The definition of moments of  $X$  is meaningful since the mappings  $X \mapsto X^q$ ,  $(X - \mu)^q$  are continuous and, therefore, measurable. Hence,  $X^q$ ,  $(X - \mu)^q$  are random variables on  $(\Omega, \mathcal{F}, P)$ .

**Definition 2.25** Let  $X = (X_1, \dots, X_d)$  be an  $\mathbb{R}^d$ -valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$  and let  $q \geq 1$  be an integer. The moments of order  $q$  of  $X$  are

$$\mu(q_1, \dots, q_d) = E \left[ \prod_{i=1}^d X_i^{q_i} \right] \quad (2.23)$$

provided the expectations in (2.23) exist and are finite, where  $q_i \geq 0$  are integers such that  $\sum_{i=1}^d q_i = q$ . The mean or the expectation of  $X$  is the vector  $E[X] = (E[X_1] = \mu(1, 0, \dots, 0), \dots, E[X_d] = \mu(0, \dots, 0, 1))$ . The  $(d, d)$ -matrix  $r = \{r_{ij} = E[X_i X_j], i, j = 1, \dots, d\}$  whose entries are moments of order  $q=2$  is called the correlation matrix of  $X$ . The formula in (2.23) with  $X_i - E[X_i]$  in place of  $X_i$ ,  $i = 1, \dots, d$ , give the central moments of order  $q$  of  $X$ . The  $(d, d)$ -matrix  $c = \{c_{ij} = E[(X_i - E[X_i])(X_j - E[X_j])], i, j = 1, \dots, d\}$  is called the covariance matrix of  $X$ . The scaled version of the covariance matrix  $\rho = \{\rho_{ij} = E[(X_i - E[X_i])(X_j - E[X_j])]/(\sigma_i \sigma_j)], i, j = 1, \dots, d\}$  is the correlation coefficient matrix, where  $\sigma_i^2 = c_{ii}$ ,  $i = 1, \dots, d$ .

The definition in (2.23) is meaningful since the mapping  $X \mapsto \prod_{i=1}^d X_i^{q_i}$  is continuous so that it is measurable implying that  $\prod_{i=1}^d X_i^{q_i}$  is a random variable defined on the same probability space as  $X$ . Similar considerations hold for the mapping  $X \rightarrow \prod_{i=1}^d (X_i - E[X_i])^{q_i}$ .

**Definition 2.26** Let  $X$  be an  $\mathbb{R}^d$ -valued random variable with finite moments of order 2. Two distinct coordinates  $X_i$  and  $X_j$  of  $X$  are said to be orthogonal if  $r_{ij} = E[X_i X_j] = 0$ . If  $c_{ij} = (X_i - E[X_i])(X_j - E[X_j]) = 0$ , then  $X_i$  and  $X_j$  are said to be uncorrelated. The correlations and covariances of  $X$  coincide if  $E[X_i] = 0$  for all  $i = 1, \dots, d$ .

*Example 2.25* Let  $X = U + iV$  be a complex-valued random variable, where  $U$  and  $V$  are real-valued random variables. If  $E[|U|] < \infty$  and  $E[|V|] < \infty$ , then  $E[X]$  exists and is equal to  $E[X] = E[U] + iE[V]$ , that is, the real and imaginary parts of  $X$  are viewed as the coordinates of a two-dimensional vector.  $\diamond$

The coordinate by coordinate definition of the expectation for random vectors extends directly to random functions by viewing their values at various arguments as coordinates of a random vector with finite, countable, or uncountable number of coordinates.

**Definition 2.27** Let  $X(t) = (X_1(t), \dots, X_d(t))$ ,  $t \in D \subseteq \mathbb{R}^{d'}$ , be an  $\mathbb{R}^d$ -valued random function defined on a probability space  $(\Omega, \mathcal{F}, P)$ . The expectation of  $X(t)$  exists if and only if the expectation of all its coordinates  $\{X_i(t), t \in D, i = 1, \dots, d\}$  exist. In this case, we have  $E[X(t)] = (E[X_1(t)], \dots, E[X_d(t)])$ ,  $t \in D$ .

*Example 2.26* Let  $Y_1$  and  $Y_2$  be random variables on a probability space  $(\Omega, \mathcal{F}, P)$  such that  $E[|Y_k|] < \infty$ ,  $k=1, 2$ . The  $\mathbb{R}^2$ -valued random function  $X(t) = (X_1(t) = Y_1 \cos(t), X_2(t) = Y_2 \sin(t))$ ,  $t \in [0, 2\pi]$ , is defined on  $(\Omega, \mathcal{F}, P)$  and has expectation  $(E[X_1(t)] = E[Y_1] \cos(t), E[X_2(t)] = E[Y_2] \sin(t))$ ,  $t \in [0, 2\pi]$ .  $\diamond$

We conclude this section with the statement of Fubini's theorem specifying conditions under which integrals of measurable functions defined on product probability

spaces can be performed sequentially. The theorem is used extensively in calculations, as illustrated by two examples.

**Theorem 2.9** (Fubini's theorem) *Let  $(\Omega_k, \mathcal{F}_k, P_k)$ ,  $k = 1, 2$ , be two complete probability spaces and denote by  $\Omega = \Omega_1 \times \Omega_2$ ,  $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$ , and  $P = P_1 \times P_2$  the product sample space,  $\sigma$ -field, and probability measure. If  $(\omega_1, \omega_2) \mapsto X(\omega_1, \omega_2)$  is  $\mathcal{F}$ -measurable and  $P$ -integrable, then*

$$\begin{aligned} X(\omega_1, \cdot) \text{ is } \mathcal{F}_2\text{---measurable and } P_2\text{---integrable for each } \omega_1 \in \Omega_1, \\ \int_{\Omega_2} X(\cdot, \omega_2) P_2(d\omega_2) \text{ is } \mathcal{F}_1\text{---measurable and } P_1\text{---integrable, and} \\ \int_{\Omega} X(\omega) P(d\omega) = \int_{\Omega_1} \left[ \int_{\Omega_2} X(\omega_1, \omega_2) P_2(d\omega_2) \right] P_1(d\omega_1). \end{aligned} \quad (2.24)$$

*If in addition  $X$  is positive and either side of (2.24) exists and is finite or infinite, so is the other side of the equality, and the equality is valid ([4], p. 59).*

**Example 2.27** Let  $(\Omega, \mathcal{F}, P)$  and  $([0, 1], \mathcal{H}([0, 1]), \lambda)$  be measure spaces, where  $\lambda$  denotes the Lebesgue measure. Let  $X : ([0, 1] \times \Omega, \mathcal{B}([0, 1]) \times \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$  be a measurable function defined on the product of these spaces endowed with the product measure  $\lambda \times P$ . It is common to interpret the first argument of  $X$  as time. The integral  $I(A, \omega) = \int_A 1_B(X(s, \omega)) ds$ ,  $A \in \mathcal{B}([0, 1])$ ,  $B \in \mathcal{B}$ , represents the time  $X(\cdot, \omega)$ ,  $\omega \in \Omega$ , spends in  $B$  during a time interval  $A$ . The expectation of this occupation time is  $E[I(A, \omega)] = \int_A P(X(s) \in B) ds$ .  $\diamond$

*Proof* The measurable mapping  $(s, \omega) \mapsto X(s, \omega)$  is said to be a stochastic process, and  $s \mapsto X(s, \omega)$  is sample  $\omega$  of  $X$ . For  $\mathcal{H} = \{\emptyset, \{0, 1\}, \{0\}, \{1\}\}$ , the indicator function,  $1_B : (\mathbb{R}, \mathcal{B}) \rightarrow (\{0, 1\}, \mathcal{H})$ ,  $B \in \mathcal{B}$ , is measurable so that  $1_B \circ X : ([0, 1] \times \Omega, \mathcal{B}([0, 1]) \times \mathcal{F}) \rightarrow (\{0, 1\}, \mathcal{H})$  is also measurable. The expectation of the occupation time is

$$\begin{aligned} E[I(A, \omega)] &= \int_{\Omega \times A} 1_B(X(s, \omega)) ds P(d\omega) = \int_{\Omega} \left[ \int_A 1_B(X(s, \omega)) ds \right] P(d\omega) \\ &= \int_A \left[ \int_{\Omega} 1_B(X(s, \omega)) P(d\omega) \right] ds = \int_A P(X(s) \in B) ds, \end{aligned}$$

by Fubini's theorem ([11], Example 5.9.1).  $\blacktriangle$

**Example 2.28** Consider a cantilever beam with unit length and stiffness that is fixed at its left end and subjected to a distributed random load  $X(x)$ ,  $x \in [0, 1]$ . It is assumed that the mapping  $(x, \omega) \mapsto X(x, \omega)$  is measurable from  $([0, 1] \times \Omega, \mathcal{B}[0, 1] \times \mathcal{F})$  to  $(\mathbb{R}, \mathcal{B})$  and  $\lambda \times P$ -integrable, where  $\lambda$  denotes the Lebesgue measure on the real line. The beam displacement  $U(1)$  at its free end and its expectations are

$$U(1) = \int_{[0,1]^3} X(x+u)u1(0 \leq u \leq 1-y, 0 \leq y \leq z) du dy dz \quad \text{and}$$

$$E[U(1)] = \int_{[0,1]^3} E[X(x+u)]u1(0 \leq u \leq 1-y, 0 \leq y \leq z) du dy dz.$$

If  $E[X(x)] = q$  is space invariant, then  $E[U(1)] = q/8$ .  $\diamond$

*Proof* The beam displacement satisfies the equation  $U''(x) = -M(x)$  with boundary conditions  $U(0) = 0$  and  $U'(0) = 0$ , where  $M(x) = -\int_0^{1-x} X(x+u)u du$  denotes the bending moment in the beam. The solution of this equation is  $U(x) = \int_0^x dz \int_0^z dy \int_0^{1-y} X(x+u)u du$ , so that

$$E[U(1)] = \int_{\Omega \times [0,1]^3} X(x+u, \omega)u1(0 \leq u \leq 1-y, 0 \leq y \leq z) du dy dz P(d\omega),$$

where  $du dy dz$  is the Lebesgue measure on  $[0, 1]^3$ . Since  $(u, y, z, \omega) \mapsto X(u, y, z, \omega)$  is measurable, we have

$$E[U(1)] = \int_{[0,1]^3} \left[ \int_{\Omega} X(x+u, y, z, \omega) P(d\omega) \right] u1(0 \leq u \leq 1-y, 0 \leq y \leq z) du dy dz,$$

by Fubini's theorem, which gives the stated formula of  $E[U(1)]$ .  $\blacktriangle$

*Example 2.29* Let  $X(t, \omega)$  and  $Y(t, \omega)$  be real-valued random variables defined on a probability space  $([0, \tau] \times \Omega, \mathcal{B}[0, \tau] \times \mathcal{F}, \lambda \times P)$ ,  $0 < \tau < \infty$ . Suppose  $X$  is the solution of the differential equation  $\ddot{X}(t, \omega) + 2\zeta v_0 \dot{X}(t, \omega) + v_0^2 X(t, \omega) = Y(t, \omega)$  with  $X(0, \omega) = 0$ ,  $\dot{X}(t, \omega) = 0$ , where  $\zeta \in (0, 1)$  and  $v_0 > 0$  are constants, and the dots denote time derivatives. We will see that the solution of this equation is a stochastic process, that is, a family of random variables indexed by  $t \in [0, \tau]$ , defined by  $X(t, \omega) = \int_0^t h(t-s)Y(s, \omega) ds$ , where  $h(u) = \exp(-\zeta v_0 u) \sin(v_d u)/v_d$  and  $v_d = v_0 \sqrt{1 - \zeta^2}$ . The expectation of this process at an arbitrary time  $t$  is the double integral  $E[X(t, \omega)] = \int_{\Omega} \left[ \int_0^t h(t-s)Y(s, \omega) ds \right] P(d\omega)$ . If  $h(t-s)Y(s, \omega)$  is  $\mathcal{B}[0, \tau] \times \mathcal{F}$ -measurable and  $\lambda \times P$ -integrable, Fubini's theorem applies and gives

$$E[X(t, \omega)] = \int_0^t h(t-s)E[Y(s, \omega)] ds \quad (2.25)$$

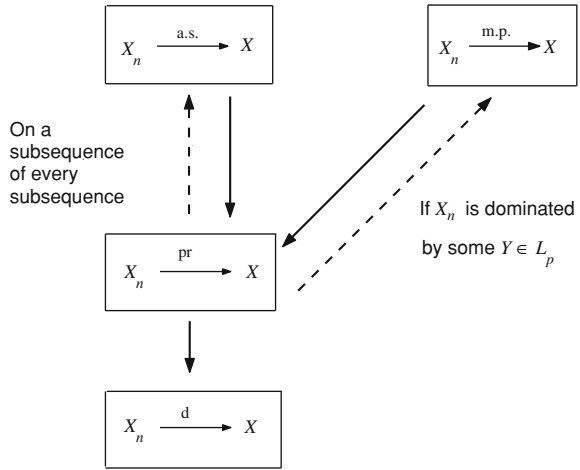
so that  $E[X(t, \omega)] = \mu_y \int_0^t h(t-s) ds$  if  $E[Y(s, \omega)] = \mu_y$  is time invariant.  $\diamond$

## 2.7 Convergence of Sequences of Random Variables

Let  $X$  and  $X_n$ ,  $n = 1, 2, \dots$ , be real-valued random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$ . There are various definitions for the convergence  $X_n \rightarrow X$  depending on the manner in which the discrepancy between  $X_n$  and  $X$  is measured.



**Fig. 2.1** Relations between the convergence of sequences of random variables



**Definition 2.28** The sequence  $\{X_n\}$  converges to  $X$  almost surely ( $X_n \xrightarrow{\text{a.s.}} X$ ), in probability ( $X_n \xrightarrow{\text{pr}} X$ ), in distribution ( $X_n \xrightarrow{\text{d}} X$ ), and in  $L^p$  ( $X_n \xrightarrow{\text{m.p.}} X$ ), where  $p \geq 1$  is an integer, if

$$\begin{aligned}
 \lim_{n \rightarrow \infty} X_n(\omega) &= X(\omega), \quad \forall \omega \in \Omega \setminus N \text{ with } P(N) = 0, \\
 \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) &= 0, \quad \forall \varepsilon > 0, \\
 \lim_{n \rightarrow \infty} F_n(x) &= F(x) \text{ at continuity points } x \in \mathbb{R}, \text{ and} \\
 \lim_{n \rightarrow \infty} E[|X_n - X|^p] &= 0,
 \end{aligned} \tag{2.26}$$

respectively. The convergence  $X_n \xrightarrow{\text{m.p.}} X$  for  $p = 2$  is called mean square (m.s.) convergence and is denoted by  $X_n \xrightarrow{\text{m.s.}} X$  or l.i.m.  $\lim_{n \rightarrow \infty} X_n = X$ .

Figure 2.1 adapted from [7] (Sect. 2.13) gives essential relations between various types of convergence. An extensive discussion on relationships between convergence types can be found, for example, in [4] (Chap. 4) and [11] (Sects. 6.3 and 8.5). We only discuss some of the arrows in Fig. 2.1. That m.s. convergence implies convergence in probability follows from the Chebyshev inequality.

The convergence  $X_n \xrightarrow{\text{a.s.}} X$  means  $P(|X_n - X| > \varepsilon \text{ i.o.}) = 0$  for  $\varepsilon > 0$  arbitrary, since it requires that  $|X_n(\omega) - X(\omega)|$  gets small and remains small for almost all  $\omega \in \Omega$ . We have  $0 = P(\limsup_{n \rightarrow \infty} |X_n - X| > \varepsilon) = \lim_{n \rightarrow \infty} P(\cup_{m \geq n} |X_m - X| > \varepsilon) \geq \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon)$ , that is,  $X_n \xrightarrow{\text{pr}} X$ .

To show that the convergence in distribution is implied by that in probability, note that for  $\varepsilon > 0$  arbitrary, we have

$$\begin{aligned}
 P(X_n \leq x) &= P(X_n \leq x, |X_n - X| \leq \varepsilon) + P(X_n \leq x, |X_n - X| > \varepsilon) \\
 &\leq P(X_n \leq x, |X_n - X| \leq \varepsilon) + P(|X_n - X| > \varepsilon)
 \end{aligned}$$

and  $P(X_n \leq x, |X_n - X| \leq \varepsilon) \leq P(X_n \leq x, X \leq X_n + \varepsilon) \leq P(X \leq x + \varepsilon)$ , which gives  $P(X_n \leq x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon)$ . This inequality and a similar inequality obtained by interchanging  $X$  with  $X_n$  yield  $P(X \leq x - \varepsilon) - P(|X_n - X| > \varepsilon) \leq P(X_n \leq x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon)$ , so that  $\lim_{n \rightarrow \infty} P(X_n \leq x) = P(X \leq x)$  provided the distribution of  $X$  is continuous at  $x$ .

**Example 2.30** Let  $\{X_n\}$  be a sequence of random variables converging in probability to  $X$ . Then  $\{X_n\}$  is a Cauchy sequence in probability, that is, for arbitrary  $\varepsilon > 0$  and  $\eta > 0$  there exists  $n(\varepsilon, \eta)$  such that  $P(|X_m - X_n| > \varepsilon) < \eta$  for  $m, n \geq n(\varepsilon, \eta)$ .  $\diamond$

*Proof* Since  $P(|X_m - X| + |X_n - X| > \varepsilon) \leq P(|X_m - X| > \varepsilon/2) + P(|X_n - X| > \varepsilon/2)$  and  $P(|X_m - X| > \varepsilon/2), P(|X_n - X| > \varepsilon/2) \rightarrow 0$  as  $m, n \rightarrow \infty$ , the sequence  $\{X_n\}$  is Cauchy in probability.  $\blacktriangle$

**Example 2.31** If  $\{X_k\}$  are uncorrelated random variables with finite mean  $\mu$  and variance  $\sigma^2$ , then  $\sum_{k=1}^n (X_k - \mu)/n \xrightarrow{\text{pr}} 0$  as  $n \rightarrow \infty$ . The convergence is referred to as the weak law of large numbers ([9], p. 36).  $\diamond$

**Example 2.32** If  $\{X_k\}$  are independent identically distributed (iid) random variables with finite expectation, then  $\sum_{k=1}^n X_k/n \xrightarrow{\text{a.s.}} E[X_1]$  as  $n \rightarrow \infty$ . The convergence is referred to as the strong law of large numbers [11] (Sects. 7.4 and 7.5). It shows that averages along almost all infinite samples of the sequence  $(X_1, X_2, \dots)$  are equal to  $E[X_1]$ .  $\diamond$

**Example 2.33** If  $\{X_k\}$  are iid random variables with finite mean  $\mu = E[X_1]$  and variance  $\sigma^2 = E[(X_1 - \mu)^2]$ , then  $(1/\sqrt{n}) \sum_{k=1}^n (X_k - \mu)/\sigma \xrightarrow{d} N(0, 1)$  as  $n \rightarrow \infty$ . The convergence is referred to as the central limit theorem ([11], Sects. 8.2).  $\diamond$

**Theorem 2.10** (Dominated, bounded, and monotone convergence theorems) *Let  $X_n, n = 1, 2, \dots$ , be real-valued random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$  such that  $\lim_{n \rightarrow \infty} X_n = X$  a.s.,  $A \in \mathcal{F}$ , and  $Y$  a random variable defined on the same probability space. If*

- (1)  $|X_n| \leq Y$  a.s.,  $Y \geq 0$ , and  $\int_A Y dP < \infty$ ,
- (2)  $|X_n| \leq c$  a.s. for a positive constant  $c$ , or
- (3)  $X_n \geq 0$  a.s. is an increasing sequence that can take on the value  $+\infty$ , then

$$\lim_{n \rightarrow \infty} \int_A X_n dP = \int_A (\lim_{n \rightarrow \infty} X_n) dP = \int_A X dP, \quad (2.27)$$

an equality referred to as the dominated convergence, the bounded convergence, or the monotone convergence under condition (1), condition (2), or condition (3), respectively ([4], Sect. 3.2).

The interchange of limit and integral operators in (2.27) resembles a property of Riemann integrals. For example, if the sequence of real-valued functions  $\{h_n(x)\}$

converges uniformly to  $h(x)$  on  $[a, b]$ , we have  $\int_a^b h(x) dx = \int_a^b \lim_{n \rightarrow \infty} h_n(x) dx = \lim_{n \rightarrow \infty} \int_a^b h_n(x) dx$ , where  $\int_a^b (\cdot) dx$  denotes a Riemann integral ([2], Theorem 30.3).

**Theorem 2.11** (Integration term by term, Fatou's lemma, Lebesgue's theorem) *If  $\{X_n\}$  and  $Y$  are random variables on a probability space  $(\Omega, \mathcal{F}, P)$  and  $A \in \mathcal{F}$ , the following three statements hold.*

*If  $\sum_n \int_A |X_n| dP < \infty$ , then  $\int_A \sum_n X_n dP = \sum_n \int_A X_n dP$ ;*

*If  $X_n \geq 0$  a.s. on  $A$ , then  $\int_A (\liminf_n X_n) dP \leq \liminf_n \int_A X_n dP$ ; and*

*If  $|X_n| \leq Y$  where  $Y \geq 0$  a.s. and  $P$  – integrable over  $A$ , then*

$$\int_A (\liminf_n X_n) dP \leq \liminf_n \int_A X_n dP \leq \limsup_n \int_A X_n dP \leq \int_A (\limsup_n X_n) dP. \quad (2.28)$$

*Proof* The proof of the first two statements in (2.28), that is, integration term by term and Fatou's lemma, can be found in [4] (Sect. 3.2). We only prove the last statement, that is, Lebesgue's theorem. Recall that  $\wedge_{k \geq n} X_k$  and  $\vee_{k \geq n} X_k$  are increasing and decreasing sequences and that  $\liminf_{n \rightarrow \infty} X_n = \sup_{n \geq 1} \inf_{k \geq n} X_k = \lim_{n \rightarrow \infty} \wedge_{k \geq n} X_k$  and  $\limsup_{n \rightarrow \infty} X_n = \inf_{n \geq 1} \sup_{k \geq n} X_k = \lim_{n \rightarrow \infty} \vee_{k \geq n} X_k$ . Note that  $\vee_{k \geq n} X_k = -\wedge_{k \geq n} (-X_k)$  so that  $\limsup_{n \rightarrow \infty} X_n = -\liminf_{n \rightarrow \infty} (-X_n)$ . We also have  $\{\vee_n X_n \leq x\} = \cap_n \{X_n \leq x\} \in \mathcal{F}$  and  $\{\wedge_n X_n > x\} = \cap_n \{X_n > x\} \in \mathcal{F}$ .

The Fatou lemma applied to sequence  $\{X_n + Y\}$ ,  $X_n + Y \geq 0$  a.s., gives

$$\int_A (\liminf_n X_n) dP + \int_A Y dP \leq \liminf_n \int_A X_n dP + \int_A Y dP$$

so that  $\int_A (\liminf_n X_n) dP \leq \liminf_n \int_A X_n dP$ . The last inequality in (2.28) results from  $\sup\{X_n, X_{n+1}, \dots\} = -\inf\{-X_n, -X_{n+1}, \dots\}$ . The middle inequality in Lebesgue's theorem is valid since  $\int_A X_n dP$  is a numerical sequence. ▲

**Example 2.34** If  $\{X_n\}$  is a sequence of real-valued random variables converging a.s. to  $X$  and there exists a random variable  $Z \geq 0$  a.s. such that  $|X_n| \leq Z$ , then  $\lim_{n \rightarrow \infty} E[X_n] = E[\lim_{n \rightarrow \infty} X_n] = E[X]$ . This convergence results by Lebesgue's theorem with  $A = \Omega$ . ◇

We conclude this section by noting that the family of random variables  $\{X\}$  defined on a probability space  $(\Omega, \mathcal{F}, P)$  with the property  $E[|X|^p] < \infty$  constitute a vector space that, with the norm  $\|X\| = (E[|X|^p])^{1/p}$  becomes a Hilbert space denoted by  $L^p(\Omega, \mathcal{F}, P)$ , where  $p \geq 1$  is an integer. Properties of  $L^p(\Omega, \mathcal{F}, P)$  relevant to our discussion are in Sect. B.5.

## 2.8 Radon–Nikodym Derivative

A direct application of Radon–Nikodym derivatives is the construction of improved Monte Carlo simulation algorithms for estimating properties of random elements.

**Definition 2.29** Let  $(\Omega, \mathcal{F})$  be a measurable space and let  $\mu, \nu: \Omega \rightarrow [0, \infty]$  be measures on this space. If  $\mu(A) = 0, A \in \mathcal{F}$ , implies  $\nu(A) = 0$ , we say that  $\nu$  is absolutely continuous with respect to  $\mu$  and indicate this property by the notation  $\nu \ll \mu$ . If  $\nu \ll \mu$  and  $\mu \ll \nu$ , then  $\nu$  and  $\mu$  are said to be equivalent measures.

*Example 2.35* Consider a measure space  $(\Omega, \mathcal{F}, \mu)$  and a measurable function  $h: (\Omega, \mathcal{F}) \rightarrow ([0, \infty), \mathcal{B}([0, \infty)))$ . Then

$$\nu(A) = \int_A h \, d\mu, \quad A \in \mathcal{F} \quad (2.29)$$

is a measure that is absolutely continuous with respect to  $\mu$ .  $\diamond$

*Proof* The set function  $\nu$  is positive by definition. It is countably additive since for  $A_n \in \mathcal{F}, n = 1, 2, \dots$ , disjoint sets, we have

$$\begin{aligned} \nu(\cup_{n=1}^{\infty} A_n) &= \int_{\cup_{n=1}^{\infty} A_n} h \, d\mu = \int \sum_{n=1}^{\infty} h 1_{A_n} \, d\mu \\ &= \sum_{n=1}^{\infty} \int h 1_{A_n} \, d\mu = \sum_{n=1}^{\infty} \int_{A_n} h \, d\mu = \sum_{n=1}^{\infty} \nu(A_n). \end{aligned}$$

The term by term integration is valid whether  $\sum_{n=1}^{\infty} \int h 1_{A_n} \, d\mu$  is or is not finite since  $h$  is positive. The measure  $\nu$  is absolutely continuous with respect to  $\mu$  since

$$\nu(A) = \int_A h \, d\mu = \int_{\Omega} h 1_A \, d\mu \leq \sup_{\omega \in A} (h(\omega)) \mu(A)$$

so that  $\mu(A) = 0$  implies  $\nu(A) = 0$  with the convention  $0 \cdot \infty = 0$ .  $\blacktriangle$

A converse of this example is provided by following theorem, referred to as the Radon–Nikodym theorem, that guarantees the existence of a measurable function  $h$  satisfying (2.29) for given measures  $\mu$  and  $\nu$  under some conditions.

**Theorem 2.12** If  $\mu$  and  $\nu$  are  $\sigma$ -finite measures on a measurable space  $(\Omega, \mathcal{F})$  such that  $\nu \ll \mu$ , then there exists a measurable function

$$h = \frac{d\nu}{d\mu}: (\Omega, \mathcal{F}) \rightarrow ([0, \infty), \mathcal{B}([0, \infty))), \quad (2.30)$$

called the Radon–Nikodym derivative of  $\nu$  with respect to  $\mu$ , such that (2.29) holds ([5], Theorem 18, p. 116).

**Example 2.36** Let  $X$  be a real-valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ . The measure  $Q$  induced on  $(\mathbb{R}, \mathcal{B})$  by  $P$ , that is, the distribution of  $X$ , is  $Q(B) = P(X^{-1}(B))$ ,  $B \in \mathcal{B}$ . Set  $F(x) = Q((-\infty, x])$  for  $B = (-\infty, x]$ ,  $x \in \mathbb{R}$ .

Suppose  $Q$  is absolutely continuous with respect to the Lebesgue measure  $\lambda$  on the real line, that is,  $\lambda(B) = 0$  implies  $Q(B) = 0$  for all  $B \in \mathcal{B}$ . Since  $Q$  is a probability measure, it is finite and, hence,  $\sigma$ -finite. Theorem 2.12 shows that there exists a Radon-Nikodym derivative  $h = dQ/d\lambda : (\mathbb{R}, \mathcal{B}) \rightarrow ([0, \infty), \mathcal{B}([0, \infty)))$  such that  $Q(B) = \int_B h d\lambda$  for all  $B \in \mathcal{B}$ . For  $B = (-\infty, x]$  we have  $F(x) = Q(B) = \int_{-\infty}^x f(u) d\lambda(u)$  with the notation  $h = f$ . The function  $f(x) = dF(x)/dx$  is called the probability density function or the density of  $X$ .  $\diamond$

## 2.9 Distribution and Density Functions

Let  $X$  be an  $\mathbb{R}^d$ -valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ . For  $u, v \in \mathbb{R}^d$ , we use the notation  $u \leq v$  ( $u < v$ ) to mean  $u_i \leq v_i$  ( $u_i < v_i$ ) for all  $i = 1, \dots, d$ . Similarly,  $(-\infty, u]$  denotes the rectangle  $\times_{i=1}^d (-\infty, u_i]$ . We have seen that the distribution  $Q$  of  $X$  is the probability measure induced on  $(\mathbb{R}^d, \mathcal{B}^d)$ , that is,  $Q(B) = P(X^{-1}(B))$ ,  $B \in \mathcal{B}^d$  (Definition 2.12). It is common to view distributions of random variables as probability measures  $Q$  defined for intervals  $B = \times_{i=1}^d (-\infty, x_i]$ . This definition of  $Q$  is not restrictive ([11], Corollary 3.2.1 and Proposition 3.2.4).

**Definition 2.30** The distribution function of  $X$  is

$$F(x) = P(X^{-1}((-\infty, x])) = P \circ X^{-1}((-\infty, x]) = Q((-\infty, x]), \quad x \in \mathbb{R}^d, \quad (2.31)$$

where  $Q = P \circ X^{-1}$  is the probability measure induced by  $P$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .

**Definition 2.31** Let  $F$  be the distribution of an  $\mathbb{R}^d$ -valued random variable that is absolutely continuous with respect to the Lebesgue measure  $\lambda$  on  $\mathbb{R}^d$ . The Radon-Nikodym derivative of  $F$  with respect to  $\lambda$ , referred to the density of  $X$ , exists and is given by

$$f(x) = \frac{\partial^d F(x)}{\partial x_1 \cdots \partial x_d}, \quad x \in \mathbb{R}^d. \quad (2.32)$$

Since  $P(\cap_{i=1}^d \{X_i \in (x_i, x_i + dx_i]\}) \simeq f(x) dx$ , the volume of  $f$  over an infinitesimal rectangle  $\times_{i=1}^d (x_i, x_i + dx_i]$  gives the probability that  $X$  takes values in this rectangle.

**Definition 2.32** The distribution of one or more coordinates of  $X$  can be obtained from the distribution or the density of  $X$ . For example, the distribution and the density of  $X_1$  are  $F_1(x_1) = F(x_1, \infty, \dots, \infty)$  and  $f_1(x_1) = \int_{\mathbb{R}^{d-1}} f(x) dx_2 \cdots dx_d$  or  $f_1(x_1) = dF_1(x_1)/dx_1$ , respectively.

Let  $X$  be a real-valued random variable. Following is a list of properties of the distribution function  $F$  of  $X$  that are useful for calculations.  $F$  is a right continuous, increasing function with range  $[0, 1]$ ;  $F$  can have only jump discontinuities and the set of these jumps is countable;  $F$  is continuous at  $x \in \mathbb{R}$  if and only if  $P(X = x) = 0$ ;  $\lim_{x \rightarrow \infty} F(x) = 1$ ;  $\lim_{x \rightarrow -\infty} F(x) = 0$ ;  $P(a < X \leq b) = F(b) - F(a) \geq 0$  for  $a \leq b$ ; and  $P(a \leq X < b) = F(b) - F(a) + P(X = a) - P(X = b)$  for  $a \leq b$ . These facts follow the properties of probability measures and real-valued functions (Exercise 2.17, [7], Sect. 2.10.1, [11], Sect. 2.1).

If the distribution  $F$  of a real-valued random variable  $X$  is absolutely continuous with respect to the Lebesgue measure on the real line, it has a density  $f(x) = dF(x)/dx$  with the following properties:  $F(b) - F(a) = \int_a^b f(x) dx$  for  $a \leq b$ ,  $f(x) = F'(x)$  so that  $\int_{-\infty}^x f(\xi) d\xi = F(x)$ ,  $f \geq 0$  since  $F$  is an increasing function, and  $\int_{-\infty}^{\infty} f(x) dx = 1$ . Note that  $f$  is not a probability measure.

*Example 2.37* Let  $X$  be a real-valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$  and let  $g : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$  be a measurable function. If  $Y = g \circ X$ , a random variable on  $(\Omega, \mathcal{F}, P)$ , has finite expectation, then

$$\begin{aligned} E[Y] &= \int_{\Omega} Y(\omega) P(d\omega) = \int_{\Omega} g(X(\omega)) P(d\omega) \\ &= \int_{\mathbb{R}} g(x) Q(dx) = \int_{\mathbb{R}} g(x) dF(x) = \int_{\mathbb{R}} g(x) f(x) dx, \end{aligned} \quad (2.33)$$

where  $Q(B) = P(X^{-1}(B))$ ,  $B \in \mathcal{B}$ , is the distribution of  $X$ .  $\diamond$

*Proof* If  $g = 1_B$ ,  $B \in \mathcal{B}$ , then  $E[Y] = P(X \in B) = Q(B)$ . If  $g = \sum_{i \in I} b_i 1_{B_i}$ , the subsets  $B_i \in \mathcal{B}$  partition  $\mathbb{R}$ ,  $I$  is a finite index set, and  $b_i$  are real constants, then  $E[Y] = \sum_{i \in I} b_i P(X \in B_i) = \sum_{i \in I} b_i Q(B_i)$  since integration is a linear operator. If  $g$  is an arbitrary positive Borel function, there exists a sequence of simple, increasing, and measurable functions  $g_n$ ,  $n = 1, 2, \dots$ , converging to  $g$  as  $n \rightarrow \infty$ . The expectations of  $g_n(X)$  calculated by all formulas in (2.33) coincide, so that (2.33) holds also for  $g(X)$  by the monotone convergence theorem. If  $g$  is an arbitrary Borel function, (2.33) holds since  $g = g^+ - g^-$  and  $\int g(X) dP = \int g^+(X) dP - \int g^-(X) dP$ .  $\blacktriangle$

*Example 2.38* A real-valued random variable  $X$  with density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right), \quad x, \mu \in \mathbb{R}, \quad \sigma > 0, \quad (2.34)$$

is said to be Gaussian with mean  $\mu$ , variance  $\sigma^2$ , and standard deviation  $\sigma$ , a property denoted by  $X \sim N(\mu, \sigma^2)$ . The function  $\phi(u) = \exp(-u^2/2)/\sqrt{2\pi}$ ,  $u \in \mathbb{R}$ , is the density of the standard Gaussian variable  $N(0, 1)$ .  $\diamond$

*Example 2.39* Consider a Cauchy random variable  $X$  with density  $f(x) = a/[\pi(a^2 + x^2)]$ ,  $x \in \mathbb{R}$ , where  $a > 0$  is a constant. The expectation of  $X$  does not exist since

$E[X^+] = E[1_{X \geq 0} X] = \int_0^\infty xf(x) dx = a \log(a^2 + x^2)/(2\pi) \big|_0^\infty = +\infty$  and  $E[X^-] = +\infty$ .  $\diamond$

**Example 2.40** The distribution of  $X = F^{-1} \circ \Phi(G)$  is  $F$ , where  $F$  is a continuous distribution,  $\Phi$  denotes the distribution of  $N(0,1)$ , that is,  $\Phi(x) = \int_{-\infty}^x \phi(u) du$ , and  $G \sim N(0, 1)$ .  $\diamond$

*Proof* Since  $X$  is a continuous function of  $G$ , it is a random variable. We have  $P(X \leq x) = P(F^{-1} \circ \Phi(G) \leq x) = P(G \leq \Phi^{-1}(F(x))) = F(x)$ ,  $x \in \mathbb{R}$ . Similar arguments show that  $\Phi(G) \sim U(0, 1)$  is a random variable uniformly distributed in  $(0,1)$ , that is,  $P(\Phi(G) \leq x) = x$ ,  $x \in [0, 1]$ . The representation  $X = F^{-1} \circ \Phi(G)$  is used to generate samples of random variables following arbitrary distributions.  $\blacktriangle$

Suppose now that  $X$  is an  $\mathbb{R}^d$ -valued random variable and  $d > 1$ . As for the case  $d = 1$ , we list the properties of the distribution function  $F$  of  $X$ :  $\lim_{x_k \rightarrow \infty} F(x)$ ,  $1 \leq k \leq d$  is the joint distribution of  $(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_d)$ ,  $\lim_{x_k \rightarrow -\infty} F(x) = 0$  for  $k \in \{1, \dots, d\}$ , function  $x_k \mapsto F(x)$  is increasing for each  $k \in \{1, \dots, d\}$ , and function  $x_k \mapsto F(x)$  is right continuous for each  $k \in \{1, \dots, d\}$ .

**Definition 2.33** Let  $X$  be an  $\mathbb{R}^d$ -valued random variable with density  $f$ . Denote by  $X^{(1)}$  and  $X^{(2)}$ , the first  $d_1 < d$  and the last  $d_2 = d - d_1$  coordinates of  $X$  for  $d \geq 2$ . The conditional density  $f^{(1|2)}$  of  $X^{(1)}$  given  $X^{(2)} = z$  is

$$f^{(1|2)}(x^{(1)} | z) = \frac{f(x^{(1)}, z)}{f^{(2)}(z)}, \quad (2.35)$$

where  $f^{(k)}$  denotes the density of  $X^{(k)}$ ,  $k = 1, 2$ , and  $x^{(1)} = (x_1, \dots, x_{d_1})$ . If  $f^{(1|2)}(x^{(1)} | z) = f^{(1)}(x^{(1)})$ , then  $X^{(1)}$  and  $X^{(2)}$  are independent.

The conditional probability  $P(A | B)$  in (2.5) provides a heuristic interpretation for the conditional density in (2.35). Since  $P(A | B) \simeq f^{(1|2)}(x^{(1)} | z) dx^{(1)}$  for  $A = \{X_1 \in (x_1, x_1 + dx_1], \dots, X_{d_1} \in (x_{d_1}, x_{d_1} + dx_{d_1}]\}$  and  $B = \{X_{d_1+1} \in (z_1, z_1 + dz_1], \dots, X_d \in (z_{d_2}, z_{d_2} + dz_{d_2}]\}$ ,  $f^{(1|2)}(x^{(1)} | z) dx^{(1)}$  represents the probability that  $X^{(1)}$  is in the infinitesimal rectangle  $(x_1, x_1 + dx_1] \times \dots \times (x_{d_1}, x_{d_1} + dx_{d_1}]$  under the condition  $X^{(2)} = z$ . A rigorous discussion on this topic can be found in [5] (Sect. 21.3, pp. 416–417).

**Definition 2.34** An  $\mathbb{R}^d$ -valued random variable  $X$  is said to be Gaussian with mean vector  $\mu$  and covariance matrix  $\gamma$  if it has the density

$$f(x) = [(2\pi)^d \det(\gamma)]^{-1/2} \exp\left[-\frac{1}{2}(x - \mu)' \gamma^{-1}(x - \mu)\right], \quad x \in \mathbb{R}^d, \quad (2.36)$$

where  $(\cdot)'$  denotes matrix transposition. We use the notation  $X \sim N(\mu, \gamma)$  to indicate that  $X$  has this property. If  $d = 2$ ,  $\mu_1 = \mu_2 = 0$ ,  $\gamma_{1,1} = \gamma_{2,2} = 1$ , and  $\gamma_{1,2} = \gamma_{2,1} = \rho$ ,  $\rho \in (-1, 1)$ , the density  $f$  is denoted by  $\phi(\cdot, \cdot; \rho)$  and has the expression

$$\phi(x_1, x_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right], \quad (x_1, x_2) \in \mathbb{R}^2, \quad (2.37)$$

and is referred to as the density of the standard bivariate Gaussian vector. The parameter  $\rho$  is the correlation coefficient between the coordinates of  $X$ .

*Example 2.41* Let  $X$  be a bivariate random vector with the density in (2.37). The conditional density of  $X_1 \mid (X_2 = z)$  is

$$f^{(1|2)}(x_1 \mid z) = \frac{\phi(x_1, z; \rho)}{\phi(z)} = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{x_1^2 - 2\rho x_1 z + z^2}{2(1-\rho^2)} + \frac{z^2}{2}\right]. \quad (2.38)$$

This density becomes  $f^{(1|2)}(x_1 \mid z) = \phi(x_1)$  for  $\rho = 0$  showing that  $X_1$  and  $X_2$  are independent for this value of  $\rho$ .  $\diamond$

*Example 2.42* Let  $X$  be an  $\mathbb{R}^d$ -valued random variable with density  $f_x$ . Let  $Y = g(X)$  where  $g: \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a measurable function defining a one-to-one mapping between  $X$  and  $Y$ . The density  $f_y$  of  $Y$  is given by

$$f_y(y) = f_x(h(y))|J|, \quad x, y \in \mathbb{R}^d, \quad (2.39)$$

where  $J = \{\partial x_i / \partial y_j, i, j = 1, \dots, d\}$  denotes the Jacobian matrix.

*Proof* Since  $x \mapsto y = g(x)$  is a one-to-one mapping,  $J$  is nonzero everywhere and so is the Jacobian of the inverse mapping  $y \mapsto x = g^{-1}(y)$ . Let  $D_x$  be a neighborhood of  $x \in \mathbb{R}^d$  and  $D_y = \{\eta \in \mathbb{R}^d : \eta = g(\xi), \xi \in D_x\}$  denote the image of  $D_x$  by the transformation  $x \mapsto y = g(x)$ . The equality  $P(X \in D_x) = P(Y \in D_y)$  can be written as  $\int_{D_x} f_x(\xi) d\xi = \int_{D_y} f_y(\eta) d\eta$ , or

$$\int_{D_y} f_x(g^{-1}(\eta)) \left| \frac{\partial(\xi_1, \dots, \xi_d)}{\partial(\eta_1, \dots, \eta_d)} \right| d\eta = \int_{D_y} f_y(\eta) d\eta$$

by a change of variables. The last equality gives (2.39).

If the mapping  $y \mapsto x = g^{-1}(y)$  has multiple solutions, we can construct a partition  $\{A_v\}$  of the  $x$ -space such that the mapping  $y \mapsto x$  is one-to-one in each  $A_v$ . The probability mass of  $D_y$  is equal to the sum  $\sum_v f_v |J_v|$  of the corresponding contributions of the subsets  $A_v$ , where each term  $f_v |J_v|$  is equal to the right side of (2.39) for the restriction of  $y \mapsto x = g^{-1}(y)$  to  $A_v$ .  $\blacktriangle$

*Example 2.43* Let  $X \sim N(0, 1)$  and  $Y = \cos(X)$ . The distribution of  $Y$  is  $P(Y \leq y) = \sum_{k \in \mathbb{Z}} (\Phi(2k\pi + \cos^{-1}(y)) - \Phi(2k\pi - \cos^{-1}(y)))$  for  $|y| \leq 1$  since  $\{Y \leq y\}$  if  $X$  belongs to  $\cup_{k \in \mathbb{Z}} [2k\pi - \cos^{-1}(y), 2k\pi + \cos^{-1}(y)]$ .  $\diamond$

*Example 2.44* Let  $X \sim N(0, \rho)$  be an  $\mathbb{R}^d$ -valued standard Gaussian variable with  $\rho_{ii} = 1$ , and set  $Y_i = F_i^{-1} \circ \Phi(X_i)$ , where  $F_i$  are continuous distributions with densities  $f_i$ ,  $i = 1, \dots, d$ . The density of  $Y = (Y_1, \dots, Y_d) \in \mathbb{R}^d$  is

$$f_y(y_1, \dots, y_d) = [(2\pi)^d \det(\rho)]^{-1/2} \exp\left(-\frac{1}{2} x' \rho^{-1} x\right) \prod_{i=1}^d \frac{f_i(y_i)}{\phi(x_i)}, \quad (2.40)$$



where  $x_i = \Phi^{-1} \circ F_i(y_i)$ ,  $i = 1, \dots, d$ . The non-Gaussian vector  $Y$  is said to be a translation random vector ([6], Sect. 3.1.1).  $\diamond$

*Proof* The equality  $P(\cap_{i=1}^d Y_i \leq y_i) = P(\cap_{i=1}^d X_i \leq x_i)$  holds since the mappings  $y_i \mapsto x_i = \Phi^{-1} \circ F_i(y_i)$  are invertible. This gives (2.40) by differentiation. Note that (2.40) follows directly from (2.39) since  $|J| = \prod_{i=1}^d f_i(y_i)/\phi(x_i)$ .  $\blacktriangle$

## 2.10 Characteristic Function

The characteristic function defines completely the probability law of random variables. We describe random variables by their distributions or characteristic functions depending on the objective of the analysis.

**Definition 2.35** Let  $X$  be an  $\mathbb{R}^d$ -valued random variable with distribution  $F$ . The characteristic function of  $X$  is

$$\varphi(u) = E[e^{iu'X}] = \int_{\mathbb{R}^d} e^{iu'x} dF(x) = \int_{\mathbb{R}^d} e^{iu'x} f(x) dx, \quad u \in \mathbb{R}^d, \quad (2.41)$$

where  $f$  is the density of  $X$ , provided it exists and  $u'$  denotes the transpose of  $u \in \mathbb{R}^d$ . The expectation of the complex-valued random variable  $\exp(iu'X)$  is obtained from the expectations of its real and imaginary parts, that is,  $E[\exp(iu'X)] = E[\cos(u'X)] + iE[\sin(u'X)]$ . The characteristic function is always defined.

Suppose first that  $X$  is a real-valued random variable. Following is a list of properties of the characteristic function of  $X$  that are useful for calculations:  $|\varphi(u)| \leq \varphi(0) = 1$  for all  $u \in \mathbb{R}$ ;  $\varphi(-u) = \varphi(u)^*$ , where  $z^*$  denotes the complex conjugate of  $z \in \mathbb{C}$ ,  $\varphi$  is positive definite (Exercise 2.23); the characteristic and the density functions are Fourier pairs, that is,

$$\varphi(u) = \int_{\mathbb{R}} e^{iux} f(x) dx \quad \text{and} \quad f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \varphi(u) du;$$

$\varphi$  is uniformly continuous in  $\mathbb{R}$ ; and, if  $X \in L^q$ , then  $\varphi \in C^q(\mathbb{R})$  and  $\varphi^{(k)}(0) = i^k E[X^k]$  for  $k = 1, \dots, q$ . Most of these properties result from the definition of  $\varphi$  but the proof of some properties requires technical arguments ([7], Sect. 2.10.3, [11], Sects. 9.2–9.5).

**Example 2.45** Consider a Cauchy random variable  $X$  with density  $f(x) = a/[\pi(a^2 + x^2)]$ ,  $a > 0$ ,  $x \in \mathbb{R}$  (Example 2.39). The characteristic function of  $X$  exists and is  $\varphi(u) = \exp(-a|u|)$ ,  $u \in \mathbb{R}$ . However,  $X$  has no mean since  $\varphi(u)$  is not differentiable at  $u = 0$ .  $\diamond$

**Example 2.46** Let  $F(x) = \sum_{k \in I} p_k 1(x_k \leq x)$  be the distribution of a real-valued random variable  $X$ , where  $I$  is a finite index set,  $p_k \geq 0$  such that  $\sum_{k \in I} p_k = 1$ , and  $\{x_k\}$  is an increasing sequence of real numbers. The characteristic function of  $X$  is  $\varphi(u) = \sum_{k \in I} p_k [\cos(ux_k) + i \sin(ux_k)]$  so that it does not vanish as  $|u| \rightarrow \infty$ .  $\diamond$

*Example 2.47* Let  $X = \sum_{k=1}^N Y_k$  be a compound Poisson random variable, where  $N$  is a Poisson variable with intensity  $\lambda > 0$  and  $Y_1, Y_2, \dots$  are iid random variables that are independent of  $N$ . The characteristic function of  $X$  is

$$\varphi(u) = \exp\left(-\lambda \int_{\mathbb{R}} (1 - e^{iu y}) dF_Y(y)\right) = \exp[-\lambda(1 - \varphi_Y(u))], \quad u \in \mathbb{R}, \quad (2.42)$$

where  $F_Y$  and  $\varphi_Y$  denote the distribution and the characteristic functions of  $Y_1$ .  $\diamond$

*Proof* We have  $\varphi(u) = E[e^{iuX} 1(N \geq 1) + e^{iuX} 1(N = 0)] = E[e^{iu \sum_{k=1}^N Y_k} 1(N \geq 1)] + P(N = 0)$  and  $E[e^{iu \sum_{k=1}^N Y_k} 1(N \geq 1)] = \sum_{k=1}^{\infty} (\varphi_Y(u))^k P(N = k)$ . This gives  $\varphi(u) = e^{-\lambda} \sum_{k=0}^{\infty} (\lambda \varphi_Y(u))^k / k! = \exp[-\lambda(1 - \varphi_Y(u))]$  since  $P(N = n) = e^{-\lambda} \lambda^n / n!$ ,  $n = 0, 1, \dots$ , is the probability of the Poisson variable  $N$ .  $\blacktriangle$

Suppose now that  $X$  is an  $\mathbb{R}^d$ -valued random variable and  $d > 1$ . Following is a list of relevant properties of the characteristic function  $\varphi$  of  $X$ :  $|\varphi(u)| \leq \varphi(0) = 1$  for all  $u \in \mathbb{R}^d$ ; the characteristic function and the density of  $X$  are Fourier pairs, that is,

$$\varphi(u) = \int_{\mathbb{R}^d} e^{iu'x} f(x) dx \quad \text{and} \quad f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-iu'x} \varphi(u) du;$$

$\varphi$  is uniformly continuous; and  $\varphi(u) = \prod_{k=1}^d \varphi_k(u_k)$  if  $X$  has independent coordinates, where  $\varphi_k(u_k) = E[\exp(iu_k X_k)]$ .

*Example 2.48* Let  $X \sim N(\mu, \gamma)$  be an  $\mathbb{R}^d$ -valued Gaussian variable with density given by (2.36). The characteristic function of  $X$  is

$$\varphi(u) = \exp\left(iu'\mu - \frac{1}{2}u'\gamma u\right), \quad u \in \mathbb{R}^d. \quad (2.43)$$

If  $\gamma$  is a diagonal matrix, that is, the coordinates of  $X$  are uncorrelated, they are also independent since  $\varphi(u) = \prod_{k=1}^d \varphi_k(u_k)$ , where  $\varphi_k(u_k) = E[\exp(iu_k X_k)] = \exp(iu_k \mu_k - \gamma_{kk} u_k^2 / 2)$  and  $X_k \sim N(\mu_k, \gamma_{kk})$ . Generally, lack of correlation does not imply independence. However, independence and lack of correlation are equivalent for Gaussian variables.  $\diamond$

*Example 2.49* Let  $X \sim N(\mu, \gamma)$  be an  $\mathbb{R}^d$ -valued Gaussian variable and set  $Y = aX + b$ , where  $a$  and  $b$  are  $(d', d)$  and  $(d', 1)$  matrices with constant entries. Then  $Y$  is an  $\mathbb{R}^{d'}$ -valued Gaussian variable with mean  $a\mu + b$  and covariance matrix  $a\gamma a'$ , that is, linear transformations of Gaussian vectors are Gaussian vectors.  $\diamond$

*Proof* The mean vector and the covariance matrix of  $Y$  can be obtained by direct calculations using the definition of  $Y$  and the linearity of the expectation operator. The characteristic function of  $Y$  is

$$E[e^{iv'Y}] = E[e^{iv'aX}] e^{iv'b} = \exp\left(iu'\mu - \frac{1}{2}u'\gamma u\right) e^{iv'b}, \quad v \in \mathbb{R}^{d'}, \quad u = a'v \in \mathbb{R}^d,$$

so that  $Y$  is an  $\mathbb{R}^{d'}$ -valued Gaussian variable with the stated properties.  $\blacktriangle$

*Example 2.50* Let  $X \sim N(\mu, \gamma)$  be an  $\mathbb{R}^d$ -valued random variable and denote the first  $d_1 < d$  and the last  $d_2 = d - d_1$  coordinates of  $X$  by  $X^{(1)}$  and  $X^{(2)}$ , respectively. The conditional vector  $\hat{X} = X^{(1)} \mid (X^{(2)} = z)$  is Gaussian with mean vector  $\hat{\mu}$  and covariance matrix  $\hat{\gamma}$  given by

$$\begin{aligned}\hat{\mu} &= \mu^{(1)} + \gamma^{(1,2)}(\gamma^{(2,2)})^{-1}(z - \mu^{(2)}) \quad \text{and} \\ \hat{\gamma} &= \gamma^{(1,1)} - \gamma^{(1,2)}(\gamma^{(2,2)})^{-1}\gamma^{(2,1)},\end{aligned}\tag{2.44}$$

where  $\mu^{(r)} = E[X^{(r)}]$  and  $\gamma^{(r,s)} = E[(X^{(r)} - \mu^{(r)})(X^{(s)} - \mu^{(s)})']$ ,  $r, s = 1, 2$ .  $\diamond$

*Example 2.51* Let  $X \sim N(\mu, \gamma)$  be a bivariate Gaussian vector. The random variable,

$$\hat{X} = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(X_2 - \mu_2),\tag{2.45}$$

is the optimal, mean square, linear estimator for  $X_1$  given  $X_2$ , where  $\sigma_1^2 = \gamma_{1,1}$ ,  $\sigma_2^2 = \gamma_{2,2}$ , and  $\rho\sigma_1\sigma_2 = \gamma_{1,2}$ .  $\diamond$

*Proof* Let  $Z = aX_2 + b$ ,  $a, b \in \mathbb{R}$ , be a linear estimator for  $X_1$ , and impose the conditions that  $Z$  is unbiased and minimizes the mean square error  $E[(Z - X_1)^2]$ . The first condition implies  $E[Z] = \mu_1$  so that  $a\mu_2 + b = \mu_1$  and  $Z = a(X_2 - \mu_2) + \mu_1$ . The mean square error

$$E[(Z - X_1)^2] = E[(a(X_2 - \mu_2) - (X_1 - \mu_1))^2] = a^2\sigma_2^2 + \sigma_1^2 - 2a\rho\sigma_1\sigma_2$$

of the estimator  $Z$  takes its minimum value at  $a = \rho\sigma_1/\sigma_2$ .  $\blacktriangle$

## 2.11 Conditional Expectation

Consider a real-valued random variable  $X$  defined on a probability space  $(\Omega, \mathcal{F}, P)$  such that  $E[|X|] < \infty$ , that is, the mean of  $X$  exists and is finite. Our objective is to define the conditional expectation  $E[X \mid \mathcal{G}]$  of  $X$  with respect to a sub- $\sigma$ -field  $\mathcal{G}$  of  $\mathcal{F}$ . The expectation  $E[X \mid \mathcal{G}]$  can be viewed as a local average of  $X$  since  $\mathcal{G}$  is coarser than  $\mathcal{F}$ .

*Example 2.52* Let  $X$  be an  $\mathbb{R}^2$ -valued Gaussian variable with mean zero and covariances  $E[X_i^2] = 1$ ,  $i = 1, 2$ , and  $E[X_1X_2] = \rho$ ,  $|\rho| < 1$ . The joint density of  $X$  and the density of  $X_1 \mid (X_2 = z)$  are given by (2.37) and (2.38). These densities show that  $X_1 \mid (X_2 = z) \sim N(\rho z, 1 - \rho^2)$  so that  $E[X_1 \mid X_2 = z] = \rho z$ , a result that can also be obtained from (2.44).  $\diamond$

*Example 2.53* The sample space, the  $\sigma$ -field  $\mathcal{F}$ , and the probability measure for the experiment of rolling two dice are  $\Omega = \{\omega = (i, j) : i, j = 1, \dots, 6\}$ , the

collection of all parts of  $\Omega$ , and  $P(\{\omega\}) = 1/36$ , respectively. Let  $X$  be a random variable defined on this space by  $X(\omega) = i + j$ . The expectation of  $X$  is  $E[X] = 7$ . Let  $\Lambda_n = \{\omega = (i, j) : i \wedge j = n\}$ ,  $n = 1, \dots, 6$ , be measurable sets partitioning  $\Omega$ , for example,  $\Lambda_4 = \{(4, 4), (4, 5), (5, 4), (4, 6), (6, 4)\}$ . The probability that an outcome  $(i, j)$  is in  $\Lambda_n$  is equal to the cardinality of  $\Lambda_n$  divided by 36. Let  $\mathcal{G} = \sigma(\Lambda_1, \dots, \Lambda_6)$  denote the  $\sigma$ -field generated by  $\{\Lambda_n, n = 1, \dots, 6\}$ , so that the members of  $\mathcal{G}$  are unions of members of  $\{\Lambda_n, n = 1, \dots, 6\}$ .

The local averages  $E[X | \Lambda_n]$  of  $X$  over  $\Lambda_n$  can be calculated simply. For example,  $E[X | \Lambda_4] = ((4 + 4) + (4 + 5) + (5 + 4) + (4 + 6) + (6 + 4))(1/5) = 46/5$  by the definition of  $X$  and the fact that the members of  $\Lambda_n$  are equally likely. Similar calculations give the other local averages, for example,  $E[X | \Lambda_6] = 12$ . In general, we have

$$E[X | \Lambda_n] = \sum_{\omega \in \Lambda_n} X(\omega) \frac{1}{\text{card}(\Lambda_n)} = \frac{1}{\text{card}(\Lambda_n)/36} \sum_{\omega \in \Lambda_n} X(\omega) \frac{1}{36} = \frac{1}{P(\Lambda_n)} \int_{\Lambda_n} X dP.$$

Hence, the conditional expectation of  $X$  with respect to  $\mathcal{G}$  is a simple random variable denoted by  $E[X | \mathcal{G}]$ , that takes the values  $\{E[X | \Lambda_n]\}$  with probabilities  $\{P(\Lambda_n)\}$ , that is, the random variable

$$E[X | \mathcal{G}] = \sum_{n=1}^6 E[X | \Lambda_n] 1_{\Lambda_n}.$$

This random variable is  $\mathcal{G}$ -measurable and has the properties  $\int_A E[X | \mathcal{G}] dP = \int_A X dP$  for all  $A \in \mathcal{G}$  and  $E\{E[X | \mathcal{G}]\} = E[X]$ .  $\diamond$

*Proof* It is obvious that  $E[X | \mathcal{G}]$  is a random variable on  $(\Omega, \mathcal{G}, P)$ . The members of  $\mathcal{G}$  are union of  $\Lambda_n$ , so that if, for example,  $A = \Lambda_k \cup \Lambda_l$ ,  $k \neq l$ , we have  $\int_A E[X | \mathcal{G}] dP = \sum_{n=1}^6 E[X | \Lambda_n] \int_{\Lambda_k \cup \Lambda_l} 1_{\Lambda_n} dP = E[X | \Lambda_k]P(\Lambda_k) + E[X | \Lambda_l]P(\Lambda_l)$ . Direct calculations give  $\int_A X dP = \int_A E[X | \mathcal{G}] dP$  for  $A \in \mathcal{G}$  arbitrary. We also have  $E\{E[X | \mathcal{G}]\} = \sum_{n=1}^6 E[X | \Lambda_n]P(\Lambda_n) = \sum_{n=1}^6 \int_{\Lambda_n} X dP = \int_{\Omega} X dP = E[X]$ .  $\diamond$

*Example 2.54* Let  $X$  and  $Y$  be real-valued random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Suppose  $Y$  is discrete taking distinct values  $y_i$ ,  $i = 1, 2, \dots$ , so that the sets  $B_i = \{Y = y_i\}$  partition  $\Omega$ . If  $P(B_i) > 0$ , the expectation of  $X$  conditional on  $Y$  is a discrete random variable denoted by  $E[X | Y]$  taking the values  $E[X | Y](\omega) = E[X | B_i]$  for  $\omega \in B_i$ , where  $E[X | B_i] = E[X | Y = y_i] = \int_{B_i} x dF(x)/P(B_i)$  and  $F$  denotes the distribution of  $X$ .  $\diamond$

*Proof* The distribution of  $X$  conditional on  $B_i$  is  $F(x | B_i) = P(X \leq x, B_i)/P(B_i)$  so that the conditional expectation of  $X$  given  $B_i$  can be calculated from  $E[X | B_i] = \int x dF(x | B_i) = \int_{B_i} x dF(x)/P(B_i)$ .

The conditional expectation  $E[X | Y]$  is equal to the local average of  $X$  over  $B_i$ . It constitutes a coarser version of  $X$  that can be viewed as an approximation of this random variable. Since the measurable partition  $\{B_i\}$  of  $\Omega$  generates the

$\sigma$ -field  $\sigma(Y)$ , we may write  $E[X | \sigma(Y)]$  for  $E[X | Y]$ . The conditional expectation  $E[X | \sigma(Y)]$  has the same properties as  $E[X | \mathcal{G}]$  in the previous example.  $\blacktriangle$

It is not possible to extend the definition of  $E[X | Y]$  in Example 2.54 to continuous random variables  $Y$  since, for example,  $\{\omega \in \Omega : a < Y(\omega) \leq b\}$ ,  $a < b$ , does not belong to the  $\sigma$ -field generated by the sets  $\{\omega \in \Omega : Y(\omega) = y\}$ ,  $y \in \mathbb{R}$ . In agreement with an observation in this example, we set  $E[X | Y]$  to be the conditional expectation  $E[X | \sigma(Y)]$ .

**Definition 2.36** Let  $X$  be a real-valued integrable random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ , and let  $\mathcal{G}$  be a sub- $\sigma$ -field of  $\mathcal{F}$ . The conditional expectation  $E[X | \mathcal{G}]$  of  $X$  with respect to  $\mathcal{G}$  is the class of  $\mathcal{G}$ -measurable functions satisfying the defining relation

$$\int_{\Lambda} X dP = \int_{\Lambda} E[X | \mathcal{G}] dP, \quad \forall \Lambda \in \mathcal{G}. \quad (2.46)$$

Note that  $E[X | \mathcal{G}]$  exists ([4], Theorem 9.1.1) and is such that  $E[X | \mathcal{G}] = E[X]$  for  $\mathcal{G} = \{\emptyset, \Omega\}$ ,  $E[X | \mathcal{G}] = X$  for  $\mathcal{G} = \mathcal{F}$  (Exercise 2.32), and  $E\{E[X | \mathcal{G}]\} = E[X]$  by (2.46) with  $\Lambda = \Omega$ .

**Theorem 2.13** Let  $X$  be a real-valued integrable random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ ,  $\mathcal{G}$  a sub- $\sigma$ -field of  $\mathcal{F}$ , and  $Z$  a real-valued  $\mathcal{G}$ -measurable function. Then

$$\begin{aligned} E[(X - E[X | \mathcal{G}])Z] &= 0, \quad \forall Z \in \mathcal{G} \\ E[XZ | \mathcal{G}] &= ZE[X | \mathcal{G}] \quad \text{a.s., } \forall Z \in \mathcal{G}. \end{aligned} \quad (2.47)$$

*Proof* The first equality in (2.47) holds for  $Z = 1_{\Delta}$ ,  $\Delta \in \mathcal{G}$ , by the defining relation. It also holds for simple random variables  $Z = \sum_n b_n 1_{\Delta_n}$ ,  $\Delta_n \in \mathcal{G}$ , since expectation is a linear operator.

The random variables  $E[XZ | \mathcal{G}]$  and  $ZE[X | \mathcal{G}]$  in the second equality of (2.47) are  $\mathcal{G}$ -measurable. If  $Z = 1_{\Delta}$ ,  $\Delta \in \mathcal{G}$ , (2.47) holds a.s., because for  $\Lambda \in \mathcal{G}$  the left and the right sides of this equation are  $\int_{\Lambda} E[X 1_{\Delta} | \mathcal{G}] dP = \int_{\Lambda} X 1_{\Delta} dP = \int_{\Lambda \cap \Delta} X dP$  and  $\int_{\Lambda} 1_{\Delta} E[X | \mathcal{G}] dP = \int_{\Lambda \cap \Delta} E[X | \mathcal{G}] dP = \int_{\Lambda \cap \Delta} X dP$ , respectively, by the defining relation. This equality also holds for simple random variables  $Z$  by the linearity of conditional expectation.

The extension of (2.47) to an arbitrary random variable  $Z$  results from the representation of  $Z$  by a difference of two positive random variables, which can be defined as limits of simple random variables ([4], Sect. 9.1).  $\blacktriangle$

**Corollary 2.1** The conditional expectation  $E[X | \mathcal{G}]$  is the projection of  $X$  on  $\mathcal{G}$  and  $X - E[X | \mathcal{G}]$  is orthogonal to  $\mathcal{G}$ . Moreover,  $E[X | \mathcal{G}]$  represents the best mean square (m.s.) estimator for  $X$  given the information content of  $\mathcal{G}$ .

*Proof* That  $X - E[X | \mathcal{G}]$  is orthogonal to  $\mathcal{G}$  follows from the first equality in (2.47). Since  $E[X | \mathcal{G}]$  is the orthogonal projection of  $X$  on  $\{Z : Z \in \mathcal{G}\}$ , it is the best m.s. estimator for  $X$  given  $\mathcal{G}$  ([8], Sects. 4.3 and 4.4).  $\blacktriangle$

**Example 2.55** The conditional expectation  $E[X \mid \sigma(Z)] = E[X \mid Z]$  is the best m.s. estimator of  $X$  with respect to the information content of  $\sigma(Z)$ , where  $X, Z \in L^2(\Omega, \mathcal{F}, P)$ . The best m.s. linear estimator of  $X$  is

$$\hat{X} = E[X] + \frac{E[XZ] - E[X]E[Z]}{E[Z^2] - E[Z]^2} (Z - E[Z]). \quad (2.48)$$

The estimator represents the linear regression of  $X$  with respect to  $Z$ , and becomes  $\hat{X} = E[X]$  if  $X$  and  $Z$  are uncorrelated.  $\diamond$

*Proof* The function  $\hat{X} = aZ + b$  is  $\sigma(Z)$ -measurable for any constants  $a, b$ . It is the conditional expectation of  $X$  with respect to  $\sigma(Z)$  if  $E[X1_\Lambda] = E[(aZ + b)1_\Lambda]$  for all  $\Lambda \in \sigma(Z)$  by (2.46), which gives  $E[X] = E[\hat{X}] = aE[Z] + b$  for  $\Lambda = \Omega$ . The orthogonality condition in (2.47) implies  $E[(X - \hat{X})Z] = 0$  or  $E[XZ] = aE[Z^2] + bE[Z]$ . The solutions  $a, b$  of these equations introduce in  $\hat{X} = aZ + b$  give the expression of  $\hat{X}$  in (2.48). The resulting estimator has the property that its m.s. error  $E[(\hat{X} - X)^2]$  is smaller than the error  $E[(aZ + b - X)^2]$  for all the other values of  $a$  and  $b$ .  $\blacktriangle$

The following three theorems give properties of the conditional expectations that are useful for calculations. The properties listed below are similar to those of expectation and hold a.s. As indicated at the beginning of this section, we consider real-valued random variables defined on the same probability space  $(\Omega, \mathcal{F}, P)$  and a sub- $\sigma$ -field  $\mathcal{G}$  of  $\mathcal{F}$ .

**Theorem 2.14** *If  $X$  and  $X_n$  are integrable random variables, then ([11], Sect. 10.3)*

$$\begin{aligned} X \in \mathcal{G} &\text{ implies } E[X \mid \mathcal{G}] = X, \\ E[aX_1 + bX_2 \mid \mathcal{G}] &= aE[X_1 \mid \mathcal{G}] + bE[X_2 \mid \mathcal{G}] \quad (\text{linearity}), \\ X_1 \leq X_2 &\text{ implies } E[X_1 \mid \mathcal{G}] \leq E[X_2 \mid \mathcal{G}] \quad (\text{monotonicity}), \\ |E[X \mid \mathcal{G}]| &\leq E[|X| \mid \mathcal{G}] \quad (\text{modulus inequality}), \\ X_n \uparrow (\downarrow) X &\text{ implies } E[X_n \mid \mathcal{G}] \uparrow (\downarrow) E[X \mid \mathcal{G}] \quad (\text{monotone convergence}), \\ |X_n| \leq Y \text{ a.s., } E[Y] < \infty \text{ a.s., and } X_n \rightarrow X &\text{ imply } E[X_n \mid \mathcal{G}] \rightarrow E[X \mid \mathcal{G}] \\ &\quad (\text{dominated convergence}). \end{aligned} \quad (2.49)$$

**Theorem 2.15** *The Cauchy-Schwarz and Jensen inequalities are, respectively, ([4], Sect. 9.1)*

$$(E[XY \mid \mathcal{G}])^2 \leq E[X^2 \mid \mathcal{G}]E[Y^2 \mid \mathcal{G}] \quad \text{and} \quad (2.50)$$

$$g(E[X \mid \mathcal{G}]) \leq E[g(X) \mid \mathcal{G}], \quad \text{where } g: \mathbb{R} \rightarrow \mathbb{R} \text{ is a convex function.} \quad (2.51)$$

**Theorem 2.16** *If  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are sub- $\sigma$ -fields of  $\mathcal{F}$  such that  $\mathcal{G}_1 \subset \mathcal{G}_2$ , then ([4], Sect. 9.1)*

$$E[X | \mathcal{G}_1] = E[X | \mathcal{G}_2] \iff E[X | \mathcal{G}_2] \in \mathcal{G}_1 \quad \text{and} \quad (2.52)$$

$$E\{E[X | \mathcal{G}_2] | \mathcal{G}_1\} = E[X | \mathcal{G}_1] = E\{E[X | \mathcal{G}_1] | \mathcal{G}_2\}. \quad (2.53)$$

**Definition 2.37** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\mathcal{G}$  a sub- $\sigma$ -field of  $\mathcal{F}$ . The conditional probability with respect to  $\mathcal{G}$  is

$$P(A | \mathcal{G}) = E[1_A | \mathcal{G}], \quad A \in \mathcal{F}. \quad (2.54)$$

The definition is meaningful since  $1_A$  is  $\mathcal{F}$ -measurable. The random variable  $P(A | \mathcal{G})$  is integrable,  $\mathcal{G}$ -measurable, and satisfies  $\int_A P(A | \mathcal{G}) dP = P(A \cap A)$  for all  $A \in \mathcal{G}$ . The latter equality holds since  $\int_A P(A | \mathcal{G}) dP = \int_A E[1_A | \mathcal{G}] dP = \int_A 1_A dP$  by the defining relation in (2.46).

*Example 2.56* Let  $A$  and  $B$  be events on a probability space  $(\Omega, \mathcal{F}, P)$  such that  $P(B) > 0$  and  $P(B^c) > 0$ , and consider the sub- $\sigma$ -field  $\mathcal{G} = \{\emptyset, \Omega, B, B^c\}$  of  $\mathcal{F}$ . The conditional probability in (2.54) is

$$\begin{aligned} P(A | \mathcal{G}) &= E[1_A | \mathcal{G}] = E[1_A | B]1_B + E[1_A | B^c]1_{B^c} \\ &= \frac{P(A \cap B)}{P(B)}1_B + \frac{P(A \cap B^c)}{P(B^c)}1_{B^c}, \end{aligned} \quad (2.55)$$

where that latter equality holds by Example 2.53. This shows that  $P(A | \mathcal{G})$  is a random variable taking the values  $P(A \cap B)/P(B)$  and  $P(A \cap B^c)/P(B^c)$  with probabilities  $P(B)$  and  $P(B^c)$ , respectively. Note that  $P(A | \mathcal{G})$  extends the definition of the conditional probability in (2.5).  $\diamond$

*Example 2.57* Let  $X \geq 0$  a.s. be a random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$  and  $\mathcal{G}$  a sub- $\sigma$ -field of  $\mathcal{F}$ . If  $E[X] < \infty$ , the set function  $Q(A) = E[X1_A]$ ,  $A \in \mathcal{G}$ , is a finite measure on the measurable space  $(\Omega, \mathcal{G})$ . The conditional expectation of  $X$  with respect to  $\mathcal{G}$  is the Radon–Nikodym derivative  $E[X | \mathcal{G}] = dQ/dP$ .

This definition becomes  $E[X | \mathcal{G}] = E[X^+ | \mathcal{G}] - E[X^- | \mathcal{G}]$  for random variables with  $E[|X|] < \infty$  and  $E[X | \mathcal{G}] = (E[X_1 | \mathcal{G}], \dots, E[X_d | \mathcal{G}])$  for  $\mathbb{R}^d$ -valued random variables with finite mean.  $\diamond$

*Proof* The defining relation for conditional expectation gives  $Q(A) = \int_A X dP = \int_A E[X | \mathcal{G}] dP$ ,  $A \in \mathcal{G}$ . Since  $Q$  and  $P$  are finite measures and  $Q$  is absolutely continuous with respect to  $P$ , the equality  $Q(A) = \int_A E[X | \mathcal{G}] dP$ ,  $A \in \mathcal{G}$ , implies  $E[X | \mathcal{G}] = dQ/dP$  by Theorem 2.12.  $\blacktriangle$

## 2.12 Discrete Time Martingales

Martingales are essentials for constructing stochastic integrals. This section provides a primer on discrete time martingales. An example is used to introduce an elementary version of stochastic integrals.

**Definition 2.38** Let  $(\Omega, \mathcal{F})$  be a measurable space. An increasing collection  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$  of sub- $\sigma$ -fields of  $\mathcal{F}$  is said to be a filtration in  $(\Omega, \mathcal{F})$ . A probability space  $(\Omega, \mathcal{F}, P)$  endowed with a filtration  $(\mathcal{F}_n)_{n \geq 0}$  is called a filtered probability space and is denoted by  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, P)$ . It is assumed that  $\mathcal{F}_0$  contains all the  $P$ -null sets of  $\mathcal{F}$ .

*Example 2.58* Suppose the sequence  $X = (X_1, X_2, \dots)$  gives outcomes of coin tosses. The information content of the  $\sigma$ -field  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$  is sufficient to decide whether an event related to the first  $n$  tosses has or has not occurred. For example, the event  $A = \{\text{at least 2 heads in the first five tosses}\}$  is  $\mathcal{F}_5$ -measurable because we can decide after five tosses whether  $A$  has or has not occurred. If  $\{\text{tail, tail, head, tail}\}$  is a sample of the first four tosses, the event  $A$  remains undecided so that  $A \notin \mathcal{F}_4$ .  $\diamond$

**Definition 2.39** Let  $(\Omega, \mathcal{F})$  and  $(\Psi, \mathcal{G})$  be measurable spaces,  $(\mathcal{F}_n)_{n \geq 0}$  a filtration on  $(\Omega, \mathcal{F})$ , and  $X = (X_0, X_1, \dots)$  a sequence of measurable functions from  $(\Omega, \mathcal{F})$  to  $(\Psi, \mathcal{G})$ . The sequence  $X$  is said to be adapted to the filtration  $(\mathcal{F}_n)_{n \geq 0}$  or  $\mathcal{F}_n$ -adapted if  $X_n$  is  $\mathcal{F}_n$ -measurable for each  $n \geq 0$ . The minimal or natural filtration of  $X = (X_0, X_1, \dots)$ , that is, the smallest  $\sigma$ -field with respect to which  $X$  is adapted, is the filtration  $\sigma(X_0, X_1, \dots, X_n)$ ,  $n \geq 0$ .

*Example 2.59* Let  $X = (X_0, X_1, \dots)$  be a real-valued sequence defined on a probability space  $(\Omega, \mathcal{F}, P)$  and set  $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$ . The sequences  $Y_n = g(X_n)$  and  $Y_n = \max_{0 \leq i \leq n} \{X_i\}$  are  $\mathcal{F}_n$ -adapted, where  $g : \mathbb{R} \rightarrow \mathbb{R}$  is a Borel measurable function.  $\diamond$

**Definition 2.40** Let  $X = (X_0, X_1, X_2, \dots)$  be a sequence of real-valued random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$  endowed with a filtration  $(\mathcal{F}_n)_{n \geq 0}$ . The sequence  $X_0, X_1, X_2, \dots$  is said to be an  $\mathcal{F}_n$ -martingale if (1)  $E[|X_n|] < \infty$ ,  $n \geq 0$ , (2)  $X$  is  $\mathcal{F}_n$ -adapted, and (3)  $E[X_n | \mathcal{F}_m] = X_m$  for  $0 \leq m \leq n$ .

If the equality in the third condition is replaced by  $\geq$  and  $\leq$ , then  $X$  is said to be an  $\mathcal{F}_n$ -submartingale and  $\mathcal{F}_n$ -supermartingale, respectively. If the random variables  $X_n$  are in  $L^p(\Omega, \mathcal{F}, P)$ ,  $X$  is called a  $p$ -integrable martingale, submartingale, or supermartingale. If  $p = 2$ , then  $X$  is said to be a square integrable martingale, submartingale, or supermartingale.

*Example 2.60* Let  $R_n = \sum_{i=1}^n X_i$ ,  $n \geq 1$ , and  $R_0 = 0$  be a random walk, where  $\{X_i\}$  are iid random variables. If the random variables  $X_i$  have finite mean,  $R = (R_0, R_1, R_2, \dots)$  is an  $\mathcal{F}_n$ -supermartingale, martingale, or submartingale depending on the sign of expectation  $E[X_1]$ , where  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ ,  $n = 1, 2, \dots$ , and  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ .  $\diamond$

*Proof* Note that  $(R_0, R_1, \dots, R_n)$  and  $(R_1, \dots, R_n)$  are  $\mathcal{F}_n$ -measurable for  $n \geq 1$  so that  $R = (R_0, R_1, R_2, \dots)$  is  $\mathcal{F}_n$ -adapted. Also, for  $n \geq m \geq 0$ , we have  $E[R_n | \mathcal{F}_m] = R_m + \sum_{i=m+1}^n E[X_i] = R_m + (n - m)E[X_1]$  so that  $R$  is a supermartingale, martingale, or submartingale if  $E[X_1]$  is negative, zero, or positive.  $\blacktriangle$



**Example 2.61** Let  $R = (R_0, R_1, \dots)$  be as in Example 2.60. If the random variables  $X_i$  have finite variance and mean zero, the sequence  $S_n = R_n^2 = \sum_{i,j=1}^n X_i X_j$ ,  $n \geq 1$ , with  $S_0 = 0$  is an  $\mathcal{F}_n$ -submartingale and  $S_n - nE[X_1^2]$  is an  $\mathcal{F}_n$ -martingale.  $\diamond$

*Proof* The sequences  $S_n$  and  $S_n - nE[X_1^2]$  have the first two defining properties for martingales. For the third property, note that

$$\begin{aligned} E[S_{n+1} | \mathcal{F}_n] &= E[(R_n + X_{n+1})^2 | \mathcal{F}_n] = E[R_n^2 + 2R_n X_{n+1} + X_{n+1}^2 | \mathcal{F}_n] \\ &= R_n^2 + 2R_n E[X_{n+1}] + E[X_{n+1}^2] = S_n + E[X_1^2] \geq S_n \end{aligned}$$

since  $R_n$  is  $\mathcal{F}_n$ -measurable,  $X_{n+1}$  is independent of  $\mathcal{F}_n$ , and  $E[X_{n+1}] = 0$ . Hence,  $S = (S_0, S_1, \dots)$  is a submartingale. Since  $E[S_{n+1} - (n+1)E[X_1^2] | \mathcal{F}_n] = (S_n + E[X_1^2]) - (n+1)E[X_1^2] = S_n - nE[X_1^2]$ ,  $S_n - nE[X_1^2]$  is a martingale.  $\blacktriangle$

Following are martingale properties resulting from their definition, Doob's decomposition, an elementary construction of stochastic integrals, two martingale inequalities, and a brief discussion on stopped martingales.

Let  $X = (X_0, X_1, \dots)$  be a sequence of random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$  with a filtration  $(\mathcal{F}_n)_{n \geq 0}$ . Then (1) if  $X$  is a submartingale, martingale, and supermartingale, its expectation is an increasing, constant, and decreasing function of time, (2)  $X$  is a martingale if it is both a submartingale and a supermartingale, (3) if  $X$  is a submartingale, then  $-X$  is a supermartingale, and (4) the third defining condition for martingales can be replaced with  $E[X_{n+1} | \mathcal{F}_n] = X_n$ ,  $n \geq 0$ .

**Theorem 2.17** Let  $X = (X_0, X_1, \dots)$  be a martingale on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, P)$ . The series  $Y = (Y_0, Y_1, \dots)$  defined by  $Y_0 = X_0 - E[X_0]$  and  $Y_n = X_n - X_{n-1}$ ,  $n \geq 1$ , is orthogonal, that is,  $E[Y_m Y_n] = 0$  for  $m \neq n$ .

*Proof* The properties of  $X$  imply that  $Y_n$  is integrable,  $\mathcal{F}_n$ -measurable, and satisfies  $E[Y_n | \mathcal{F}_m] = 0$ ,  $n > m$ . For  $n > m$ , we have  $E[Y_n Y_m] = E\{E[Y_n Y_m | \mathcal{F}_m]\} = E\{Y_m E[Y_n | \mathcal{F}_m]\} = 0$  since  $Y_m$  is  $\mathcal{F}_m$ -measurable and  $E[Y_n | \mathcal{F}_m] = 0$ .  $\blacktriangle$

**Theorem 2.18** (Doob decomposition) If  $X = (X_0, X_1, \dots)$  is an  $\mathcal{F}_n$ -submartingale, then there is an  $\mathcal{F}_n$ -martingale  $M_n$  and an increasing process  $A_n$  with the properties  $A_0 = 0$  and  $A_n \in \mathcal{F}_{n-1}$ ,  $n \geq 1$ , such that the representation

$$X_n = A_n + M_n, \quad n \geq 0, \quad (2.56)$$

holds and is unique. The representation shows that submartingales have a predictable part  $A_n$  that can be told ahead of time and an unpredictable part  $M_n$ .

*Proof* We first show that (2.56) is unique provided it exists. Note that  $E[X_n | \mathcal{F}_{n-1}] = E[A_n | \mathcal{F}_{n-1}] + E[M_n | \mathcal{F}_{n-1}] = A_n + M_{n-1}$  for  $n \geq 1$ . Substituting  $M_{n-1}$  in this equation with its expression from (2.56), we obtain the recurrence formula  $A_n = A_{n-1} + E[X_n | \mathcal{F}_{n-1}] - X_{n-1}$ , which defines  $A$  uniquely since  $A_0 = 0$ .

We now show that the decomposition in (2.56) exists, that is, that there are processes  $A$  and  $M$  with the stated properties. Let  $A_n$ ,  $n = 0, 1, \dots$ , be defined by

the above recurrence formula with  $A_0 = 0$ . Note that  $A_n \in \mathcal{F}_{n-1}$  and  $A_n \geq A_{n-1}$  since  $X_n$  is a submartingale, so that  $A_n$ ,  $n = 0, 1, \dots$ , has the stated properties. We also have

$$\begin{aligned} E[M_n | \mathcal{F}_{n-1}] &= E[X_n - A_n | \mathcal{F}_{n-1}] = E[X_n | \mathcal{F}_{n-1}] \\ &\quad - E[A_{n-1} + E[X_n | \mathcal{F}_{n-1}] - X_{n-1} | \mathcal{F}_{n-1}] = -A_{n-1} + X_{n-1} = M_{n-1} \end{aligned}$$

so that  $M$  is an  $\mathcal{F}_n$ -martingale.  $\blacktriangle$

*Example 2.62* Let  $X_n$ ,  $n \geq 0$ , be a square integrable martingale. Then  $X_n^2$  is a submartingale that admits the Doob decomposition in (2.56) with  $M_n = X_n^2 - A_n$  and  $A_n = \sum_{i=1}^n E[X_i^2 - X_{i-1}^2 | \mathcal{F}_{i-1}]$ .  $\diamond$

*Proof* The process  $X_n^2$  satisfies the first two defining conditions for martingales, and

$$\begin{aligned} E[X_n^2 | \mathcal{F}_{n-1}] &= E[(X_n - X_{n-1})^2 + 2X_n X_{n-1} - X_{n-1}^2 | \mathcal{F}_{n-1}] \\ &= E[(X_n - X_{n-1})^2 | \mathcal{F}_{n-1}] + X_{n-1}^2 \geq X_{n-1}^2 \end{aligned}$$

since  $E[(X_n - X_{n-1})^2 | \mathcal{F}_{n-1}] \geq 0$ ,  $X_{n-1}$  is  $\mathcal{F}_{n-1}$ -measurable, and  $X_n$  is a martingale. Hence,  $X_n^2$  is a submartingale, so that (2.56) holds with  $\{A_n\}$  given by  $A_n = A_{n-1} + E[X_n^2 | \mathcal{F}_{n-1}] - X_{n-1}^2$ ,  $n \geq 1$ , and  $A_0 = 0$ .  $\blacktriangle$

*Example 2.63* Let  $X_n$  denote our fortune after  $n$  rounds of a game with unit stake. Suppose  $m \geq 1$  rounds have been completed in this game, so that  $X_n - X_m$ ,  $n > m$ , gives our net total winnings/losses in the future  $m+1, \dots, n$  rounds. The best m.s. estimator of  $X_n - X_m$ ,  $n > m$ , given our knowledge  $\mathcal{F}_m$  after  $m$  rounds is the conditional expectation  $E[X_n - X_m | \mathcal{F}_m]$ , where  $\mathcal{F}_m = \sigma(X_1, \dots, X_m)$ ,  $m \geq 1$ , and  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ . If  $X_0, X_1, \dots$  is a martingale, then  $E[X_n - X_m | \mathcal{F}_m] = 0$ , that is, our average fortune  $E[X_n | \mathcal{F}_m]$  at time  $n > m$  is equal to our current fortune  $X_m$ .

Suppose now that stakes  $A_i$ ,  $i = 0, 1, \dots$ , other than one are allowed, where  $A_0 = 0$ . Since stakes for round  $m+1$  are decided based on knowledge  $\mathcal{F}_m$  accumulated after  $m$  rounds,  $A_{m+1}$  is  $\mathcal{F}_m$ -measurable. Processes with this property are said to be predictable processes. The sequence  $M = (M_0, M_1, M_2, \dots)$  defined by

$$M_n = \sum_{i=1}^n A_i(X_i - X_{i-1}), \quad n = 1, 2, \dots, \quad (2.57)$$

with  $M_0 = 0$  and  $X_0 = 0$  gives our total fortune after  $n \geq 1$  rounds and constitutes a discrete version of the stochastic integral studied later in the book (Chap. 4). The integrand  $\{A_i\}$  is a predictable process and the integrator is defined by increments  $\{X_i - X_{i-1}\}$  of a martingale.  $\diamond$

**Theorem 2.19** Let  $X = (X_0, X_1, \dots)$  be a square integrable  $\mathcal{F}_n$ -martingale and  $A = (A_0, A_1, \dots)$  be an  $\mathcal{F}_n$ -predictable process such that  $A_0 = 0$  and  $E[A_n^2] < \infty$ . Then  $M_n$  in (2.57) is an  $\mathcal{F}_n$ -martingale.

*Proof* Note that  $E[|M_n|] \leq \sum_{i=1}^n E[|A_i||X_i - X_{i-1}|] \leq \sum_{i=1}^n (E[A_i^2]E[(X_i - X_{i-1})^2])^{1/2}$  by the Cauchy–Schwarz inequality, so that  $E[|M_n|] < \infty$  since  $A_i$  and  $X_i$  have finite second moments. That  $M_n \in \mathcal{F}_n$  follows from its definition and the properties of  $A_n$  and  $X_n$ . For  $n > m$ , we have

$$\begin{aligned} E[M_n | \mathcal{F}_m] &= E\left[M_m + \sum_{i=m+1}^n A_i(X_i - X_{i-1}) \mid \mathcal{F}_m\right] \\ &= M_m + \sum_{i=m+1}^n E[A_i(X_i - X_{i-1}) \mid \mathcal{F}_m] \\ &= M_m + \sum_{i=m+1}^n E\{E[A_i(X_i - X_{i-1}) \mid \mathcal{F}_{i-1}] \mid \mathcal{F}_m\} = M_m \end{aligned}$$

since  $A_i \in \mathcal{F}_{i-1}$  and  $X_i$  is a martingale so that  $E[A_i(X_i - X_{i-1}) \mid \mathcal{F}_{i-1}] = A_i E[X_i - X_{i-1} \mid \mathcal{F}_{i-1}] = 0$ .  $\blacktriangle$

**Definition 2.41** An  $\{0, 1, \dots\}$ -valued random variable  $T$  defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 1}, P)$  is a stopping time with respect to  $\mathcal{F}_n$ ,  $n = 0, 1, \dots$ , or an  $\mathcal{F}_n$ -stopping time if  $\{T \leq n\} \in \mathcal{F}_n$  for each  $n \geq 0$ .

Stopping times are useful for both applications and theoretical considerations. For example, suppose  $\{X_n\}$  is the state of a physical system that performs according to specifications as long as its state does not exceed a critical value  $x_{\text{cr}}$ . The failure time  $T = \min\{n : X_n > x_{\text{cr}}\}$  is a stopping time. Stopping times are also useful tools for constructing stochastic integrals (Sect. 4.4.3). Useful information on stopping times can be found in [1] (Sect. 2.2), [7] (Sect. 2.16), [10] (Sect. 1.5), and [13] (Sect. 2.2).

**Theorem 2.20**  $T$  is a stopping time if and only if  $\{T = n\} \in \mathcal{F}_n$  for all  $n \geq 0$ .

*Proof* If  $T$  is a stopping time, then  $\{T \leq n\} \in \mathcal{F}_n$  and  $\{T \leq n-1\}^c \in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n$  so that  $\{T = n\} = \{T \leq n\} \cap \{T \leq n-1\}^c \in \mathcal{F}_n$ . If  $\{T = n\} \in \mathcal{F}_n$  for each  $n \geq 0$ , then  $\{T = m\} \in \mathcal{F}_m \subseteq \mathcal{F}_n$  for  $m \leq n$  and  $\{T \leq n\} = \bigcup_{m=0}^n \{T = m\} \in \mathcal{F}_n$ .  $\blacktriangle$

**Definition 2.42** Let  $X = (X_0, X_1, \dots)$  be an  $\mathcal{F}_n$ -submartingale, martingale, or supermartingale and let  $T$  denote an  $\mathcal{F}_n$ -stopping time. Then  $X_n^T(\omega) = X_{n \wedge T(\omega)}(\omega)$ ,  $n = 0, 1, \dots$ , is called the sequence  $X$  stopped at  $T$ . Note that the samples of  $\{X_n^T\}$  are constant at times exceeding  $T$ .

**Theorem 2.21** If  $T$  is an  $\mathcal{F}_n$ -stopping time and  $X_n$  is an  $\mathcal{F}_n$ -submartingale, martingale, and supermartingale so is  $X_n^T$ .

*Proof* Since

$$E[|X_{n \wedge T}|] = \sum_{k=0}^n \int_{\{T=k\}} |X_k| dP + \int_{\{T>n\}} |X_n| dP,$$

$E[|X_n|] < \infty$ ,  $\int_{\{T=k\}} |X_k| dP \leq E[|X_k|]$ , and  $\int_{\{T>n\}} |X_n| dP \leq E[|X_n|]$ ,  $X_n^T$  is integrable.  $X_{n \wedge T}$  is  $\mathcal{F}_n$ -measurable for all  $n \geq 0$  since  $X_{n \wedge T} = \sum_{k=0}^{n-1} X_k 1(T = k) + X_n 1(T \geq n)$ ,  $X_k \in \mathcal{F}_k \subseteq \mathcal{F}_n$  for  $k \leq n$ ,  $1(T = k) \in \mathcal{F}_k \subseteq \mathcal{F}_n$ , and  $1(T \geq n) \in \mathcal{F}_n$ . The representation  $X_{n \wedge T} = \sum_{k=0}^{n-1} X_k 1(T = k) + X_n 1(T \geq n)$  implies  $X_{(n+1) \wedge T} - X_{n \wedge T} = (X_{n+1} - X_n) 1(T > n)$  so that  $E[X_{(n+1) \wedge T} - X_{n \wedge T} | \mathcal{F}_n] = 1(T > n) E[X_{n+1} - X_n | \mathcal{F}_n]$  since  $1(T > n)$  is  $\mathcal{F}_n$ -measurable. If  $X$  is a submartingale, martingale, or supermartingale,  $E[X_{n+1} - X_n | \mathcal{F}_n]$  is positive, zero, or negative, that is,  $X_n^T$  is a submartingale, martingale, or supermartingale, respectively.  $\blacktriangle$

**Theorem 2.22** (Optional stopping theorem) *If (1)  $X$  is an  $\mathcal{F}_n$ -martingale, (2)  $T$  is a stopping time with respect to  $\mathcal{F}_n$  such that  $T < \infty$  a.s., (3)  $X_T$  is integrable, and (4)  $E[X_{n+1} 1(T > n)] \rightarrow 0$  as  $n \rightarrow \infty$ , then  $E[X_T] = \mu$ , where  $\mu = E[X_0]$ .*

*Proof* Since  $X_T = X_{n \wedge T} + (X_T - X_n) 1(T > n)$  and  $X$  is a martingale, we have  $E[X_T] = \mu + E[X_T 1(T > n)] - E[X_n 1(T > n)]$ . The expectation  $E[X_n 1(T > n)]$  converges to zero as  $n \rightarrow \infty$  by hypothesis. The expectation  $E[X_T 1(T > n)] = \sum_{k=n+1}^{\infty} E[X_k 1(T = k)]$  also converges to zero as  $n \rightarrow \infty$  since  $|E[X_T 1(T > n)]| \leq E[|X_T|]$  and  $X_T$  is integrable by assumption so that the series  $\sum_{k=0}^{\infty} E[X_k 1(T = k)]$  is convergent. Hence, the expectation of  $X_T$  is  $\mu$ .  $\blacktriangle$

We conclude with two inequalities that are useful in applications. The proof of these and other inequalities can be found in, for example, [5] (Chap 24) and [11] (Chap. 10).

**Theorem 2.23** (Doob maximal inequality) *If  $X = (X_0, X_1, \dots)$  is a positive  $\mathcal{F}_n$ -submartingale and  $\lambda > 0$  is an arbitrary constant, then*

$$P\left(\max_{0 \leq k \leq n} X_k \geq \lambda\right) \leq \frac{1}{\lambda} E\left[X_n 1\left(\max_{0 \leq k \leq n} X_k \geq \lambda\right)\right]. \quad (2.58)$$

**Theorem 2.24** (Doob maximal  $L^2$  inequality) *If  $X = (X_0, X_1, \dots)$  is a square integrable positive  $\mathcal{F}_n$ -submartingale, then*

$$E\left[\left(\max_{0 \leq k \leq n} X_k\right)^2\right] \leq 4E[X_n^2]. \quad (2.59)$$

## 2.13 Monte Carlo Simulation

Let  $X$  be an  $\mathbb{R}^d$ -valued random variable with distribution  $F$  that is defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Our objectives are to generate independent samples of  $X$  and estimate properties of  $X$  from its samples.

### 2.13.1 Gaussian Variables

Let  $X \sim N(\mu, \gamma)$  be an  $\mathbb{R}^d$ -valued Gaussian variable with mean  $\mu$  and covariance matrix  $\gamma$ . A useful representation of  $X$  is provided by the Cholesky decomposition showing that

$$X \stackrel{d}{=} \mu + \beta G \sim N(\mu, \gamma), \quad (2.60)$$

where  $G$  is an  $\mathbb{R}^d$ -valued random variable with independent  $N(0,1)$  coordinates and  $\beta$  is a lower triangular matrix whose non-zero entries are

$$\beta_{ij} = \frac{\gamma_{ij} - \sum_{r=1}^{j-1} \beta_{ir} \beta_{jr}}{\left[ \gamma_{jj} - \sum_{r=1}^{j-1} \beta_{jr}^2 \right]^{1/2}}, \quad 1 \leq j \leq i \leq d, \quad (2.61)$$

with the convention  $\sum_{r=1}^0 \beta_{ir} \beta_{jr} = 0$ .

Samples of  $X$  can be calculated from (2.60) in which  $G$  is replaced with samples of this vector generated by, for example, the MATLAB function `randn`. Alternatively, algorithms using memoryless transformations of some random variables can be used to generate independent samples of  $N(0, 1)$  variables. For example,  $Z_1 = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2)$  and  $Z_2 = \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$  are independent  $N(0,1)$  variables, where  $U_1, U_2 \sim U(0, 1)$  are independent random variables distributed uniformly in  $(0,1)$  [3, 6, 12].

### 2.13.2 Non-Gaussian Variables

Let  $X$  be a real-valued, non-Gaussian random variable with distribution  $F$  that is continuous. Samples of  $X$  can be calculated from samples of  $U(0,1)$  by the following transformation.

**Theorem 2.25** *If  $X$  is a real-valued random variable with continuous distribution  $F$ , then*

$$X \stackrel{d}{=} F^{-1}(U(0, 1)). \quad (2.62)$$

*Proof* Since  $P(F^{-1}(U(0, 1)) \leq z) = P(U(0, 1) \leq F(z)) = F(z)$ , (2.62) holds, so that samples of  $X$  can be calculated from samples of  $U(0,1)$  and the representation of  $X$  in (2.62). For example,  $n$  independent samples of an exponential random variable with mean  $1/\lambda$ ,  $\lambda > 0$ , can be obtained from  $-\ln(1 - \text{rand}(n, 1))/\lambda \stackrel{d}{=} -\ln(\text{rand}(n, 1))/\lambda$ , where `rand` is a MATLAB function generating samples of  $U(0,1)$ . ▲

Suppose now that  $X$  is an  $\mathbb{R}^d$ -valued random variable with continuous distribution  $F$ . Let  $F_1$  and  $F_{k|k-1,\dots,1}$ ,  $k = 2, \dots, d$  denote the distributions of the coordinate  $X_1$  of  $X$  and of the conditional random variable  $X_k | (X_{k-1}, \dots, X_1)$ , respectively.

**Theorem 2.26** Let  $Z = (Z_1, \dots, Z_d)$  be an  $\mathbb{R}^d$ -valued random variable defined by

$$\begin{aligned} F_1(Z_1) &= U_1, \\ F_{k|k-1,\dots,1}(Z_k | Z_{k-1}, \dots, Z_1) &= U_k, \quad k = 2, \dots, d, \end{aligned} \quad (2.63)$$

where  $\{U_k, k = 1, \dots, d\}$  are independent  $U(0,1)$  random variables. Then,  $X$  and  $Z$  have the same distribution.

*Proof* That  $P(Z_1 \leq z_1) = F_1(z_1)$  follows from Theorem 2.25. Since  $Z_2 | Z_1 \stackrel{d}{=} F_{2|1}^{-1}(U_2 | Z_1)$ , we have

$$\begin{aligned} P(Z_2 \leq z_2 | Z_1 = z_1) &= P(F_{2|1}^{-1}(U_2 | z_1) \leq z_2) \\ &= P(U_2 \leq F_{2|1}(z_2 | z_1)) = F_{2|1}(z_2 | z_1), \end{aligned}$$

and so on.  $\blacktriangle$

Let  $(u_1, \dots, u_d)$  be a sample of  $(U_1, \dots, U_d)$ . The corresponding sample  $(z_1, \dots, z_d)$  of  $Z$  can be calculated from (2.63) sequentially beginning with  $z_1 = F_1^{-1}(u_1)$  and continuing with  $F_{k|k-1,\dots,1}(z_k | z_{k-1}, \dots, z_1) = u_k$  for increasing values of  $k \geq 2$ .

*Example 2.64* Let  $X = (X_1, X_2)$  be a non-Gaussian vector with  $X_1 \sim N(\mu, \sigma^2)$  and  $X_2 | (X_1 = x_1) \sim N(x_1, \beta^2)$ . The density of  $X$  is

$$f(x_1, x_2) = \frac{1}{\sigma\beta} \phi\left(\frac{x_1 - \mu}{\sigma}\right) \phi\left(\frac{x_2 - x_1}{\beta}\right).$$

The mapping in (2.63) becomes  $Z_1 = \mu + \sigma\Phi^{-1}(U_1)$  and  $Z_2 | (Z_1 = z_1) = z_1 + \beta\Phi^{-1}(U_2)$ , where  $U_1$  and  $U_2$  are independent copies of  $U(0,1)$ .  $\diamond$

*Example 2.65* Let  $X \in \mathbb{R}^d$  be a translation vector, that is,  $X = g(Y)$ , where  $Y$  is an  $\mathbb{R}^d$ -valued Gaussian vector with mean zero, covariance matrix  $\rho = \{\rho_{i,j} = E[Y_i Y_j]\}$  such that  $\rho_{i,i} = 1$ ,  $i = 1, \dots, d$ , and  $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is Borel measurable. Samples of  $X$  can be generated from samples of  $Y$  and the definition of  $X$  or from samples of  $U(0, 1)$ , the distribution of  $X$ , and the algorithm in (2.63). The latter approach is less efficient for translation random vectors. If the mapping  $Y \mapsto X$  is given by  $X_i = F_i^{-1}(\Phi(Y_i)) = g_i(Y_i)$ ,  $i = 1, \dots, d$ , where  $F_i$  are continuous distributions, then the coordinates of  $X$  have the distributions  $F_i$ .  $\diamond$

*Proof* Let  $y_i = g_i^{-1}(x_i)$  and  $y = (y_1, \dots, y_d)$ . The distribution,

$$P(X_1 \leq x_1, \dots, X_d \leq x_d) = P(Y_1 \leq y_1, \dots, Y_d \leq y_d) = \Phi_d(y; \rho),$$

called multivariate translation distribution, can be used as input to the Monte Carlo simulation algorithm based on (2.63), where  $\Phi_d(\cdot; \rho)$  denotes the joint distribution function of the Gaussian vector  $Y \sim N(0, \rho)$ . If  $X_i = F_i^{-1}(\Phi(Y_i))$ , then  $P(X_i \leq x_i) = P(Y_i \leq \Phi^{-1}(F_i(x_i))) = F_i(x_i)$ ,  $i = 1, \dots, d$ . Details on translation random variables can be found in [6] (Sect. 3.1.1).  $\blacktriangle$

### 2.13.3 Estimators

Let  $X$  be a real-valued random variable with distribution  $F$  and  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function. Our objective is to estimate the expectation  $E[h(X)]$  from  $n$  independent samples of  $X$ . The expectation  $E[h(X)]$  is of interest in applications since it provides useful information on  $X$ . For example, if  $h(X) = X^r$  and  $r \geq 1$  is an integer, then  $E[h(X)]$  is the moment of order  $r$  of  $X$ . If  $h(X) = 1(X > a)$ ,  $a \in \mathbb{R}$ , then  $E[h(X)] = E[1(X > a)] = P(X > a)$ .

**Theorem 2.27** *Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function and let  $X_1, \dots, X_n$  be  $n$  independent copies of  $X$  such that  $E[h(X)^2] < \infty$ . The estimator,*

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n h(X_i), \quad (2.64)$$

*is unbiased, that is,  $E[\hat{Y}] = E[Y]$ , and  $\text{Var}[\hat{Y}] \rightarrow 0$  as  $n \rightarrow \infty$ , where  $Y = h(X)$ .*

*Proof* We have  $E[\hat{Y}] = (1/n) \sum_{i=1}^n E[h(X_i)] = E[Y]$  since  $X_i$  have the same distribution. Also,

$$E[\hat{Y}^2] = \frac{1}{n^2} \left[ nE[h(X_1)^2] + (n^2 - n)(E[h(X_1)])^2 \right]$$

so that  $\text{Var}[\hat{Y}] = \text{Var}[Y]/n = \text{Var}[h(X_1)]/n$ . The coefficient of variation of estimator  $\hat{Y}$  is  $\text{cov}[Y] = (\text{Var}[Y])^{1/2}/E[Y] = \text{cov}[h(X_1)]/\sqrt{n}$ .  $\blacktriangle$

The estimator  $\hat{Y}$  is guaranteed to approximate  $E[Y] = E[h(X)]$  accurately for a sufficiently large  $n$ . Yet, the required sample size  $n$  may be so large that the use of  $\hat{Y}$  becomes impractical. For example, suppose our objective is to estimate the probability  $P(X > a)$ ,  $X \sim N(0, 1)$ , that is, the expectation  $E[1(X > a)]$ . The mean and variance of  $\hat{Y}$  are  $P(X > a)$  and  $P(X > a)P(X \leq a)/n$ , respectively, so that  $\text{cov}[\hat{Y}] = \sqrt{P(X \leq a)/(nP(X > a))}$ . For  $a = 5$  we have  $E[\hat{Y}] = E[1(X > a)] = \Phi(-5) = 2.8665 \times 10^{-7}$ ,  $\text{Var}[\hat{Y}] \simeq \Phi(-5)/n$ , and  $\text{cov}[\hat{Y}] \simeq 1/\sqrt{n\Phi(-5)}$ . To have a coefficient of variation of 0.1 we need at least  $10^5$  samples. A much larger sample size would be needed for a threshold  $a > 5$ .

More efficient Monte Carlo algorithms, referred here to as improved Monte Carlo algorithms, can be constructed by measure change. Let  $P$  and  $Q$  be two probabilities on a measurable space  $(\Omega, \mathcal{F})$  such that  $P \ll Q$ , that is,  $P$  is absolutely continuous

with respect to  $Q$  (Definition 2.29). Then, there exists a positive measurable function  $g = dP/dQ : (\Omega, \mathcal{F}) \rightarrow ([0, \infty), \mathcal{B}([0, \infty)))$ , called the Radon–Nikodym derivative, such that  $P(A) = \int_A g(\omega) Q(d\omega)$ ,  $A \in \mathcal{F}$  (Sect. 2.8). Following are examples illustrating the construction of Monte Carlo simulation algorithms based on the Radon–Nikodym derivative.

*Example 2.66* Let  $X = \sum_{i=k}^m a_k 1_{A_k}$  be a simple random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ , where  $\{A_k \in \mathcal{F}, k = 1, \dots, m\}$  is a measurable partition of  $\Omega$  such that  $P(A_k) > 0$  and  $|a_k| < \infty$ ,  $k = 1, \dots, m$ . Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function. The expectation of  $h(X)$  with respect to the probability measure  $P$  is  $E_P[h(X)] = \sum_{k=1}^m h(a_k) P(A_k)$ , for example,  $E_P[h(X)] = P(X > a)$  for  $h(x) = 1(x > a)$ .

Consider another probability measure  $Q$  on the measurable space  $(\Omega, \mathcal{F})$  such that  $Q(A_k) > 0$ ,  $k = 1, \dots, m$ . We have

$$E_P[h(X)] = \sum_{k=1}^m h(a_k) P(A_k) = \sum_{k=1}^m \left[ h(a_k) \frac{P(A_k)}{Q(A_k)} \right] Q(A_k), \quad (2.65)$$

that is,  $E_P[h(X)]$  can be calculated as the expectation of random variable  $\tilde{X} = \sum_{k=1}^m h(a_k) (P(A_k)/Q(A_k)) 1_{A_k}$  with respect to the probability measure  $Q$ .  $\diamond$

*Example 2.67* Suppose our objective is to estimate the probability  $P(X > a)$  by Monte Carlo simulation, where  $X$  is a real-valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Let

$$\hat{p}_{MC}(a) = \frac{1}{n} \sum_{i=1}^n 1(x_i > a) \quad \text{and} \quad \hat{p}_{IMC}(a) = \frac{1}{n} \sum_{i=1}^n 1(z_i > a) \frac{f(z_i)}{q(z_i)},$$

be estimates of  $P(X > a)$  by direct and improved Monte Carlo simulation, where  $x_i$  and  $z_i$  are independent samples generated from the densities  $f$  and  $q$ , respectively, where  $f(\xi) = dP(X \leq \xi)/d\xi$ ,  $q(\xi) = dQ(X \leq \xi)/d\xi$ , and  $Q$  is a measure on  $(\Omega, \mathcal{F})$  such that  $P \ll Q$ .

If  $X = \exp(Y)$ ,  $Y \sim N(1, (0.2)^2)$ , then  $P(X > x)$  is 0.3712, 0.0211, 0.1603  $\times 10^{-3}$ , 0.3293  $\times 10^{-4}$ , and 0.7061  $\times 10^{-5}$  for  $a = 3, 5, 8, 9$ , and 10, respectively. The corresponding estimates  $\hat{p}_{MC}(x)$  based on 10000 samples are 0.3766, 0.0235, 0.2  $\times 10^{-3}$ , 0, and 0. The estimates  $\hat{p}_{IMC}(x)$  based on the same number of samples are 0.3733, 0.0212, 0.1668  $\times 10^{-3}$ , 0.3304  $\times 10^{-4}$ , and 0.7084  $\times 10^{-5}$  for  $a = 3, 5, 8, 9$ , and 10, where  $q(z) = \phi((z - a)/\sigma)/\sigma$  is the density of a Gaussian variable with mean  $a$  and variance  $\sigma^2$ . While the estimators  $\hat{p}_{IMC}(a)$  are satisfactory up to  $a = 10$ , the estimates  $\hat{p}_{MC}(a)$  are inaccurate for  $a \geq 8$ .  $\diamond$

*Proof* The required probability is  $P(X > a) = \int_{\mathbb{R}} 1(\xi > a) f(\xi) d\xi = E_P[1(X > a)]$ , where  $E_P$  denotes the expectation operator under  $P$ . We also have

$$P(X > a) = \int_{\mathbb{R}} \left[ 1(\xi > a) \frac{f(\xi)}{q(\xi)} \right] q(\xi) d\xi = E_Q \left[ 1(X > a) \frac{f(X)}{q(X)} \right].$$



The density  $q$  has been selected such that 50% of its samples exceed  $a$  and the ratio  $f/q$  is bounded in  $\mathbb{R}$ . ▲

*Example 2.68* Suppose our objective is to estimate the probability  $P(X > a)$ , where  $X \sim N(0, 1)$ . The coefficient of variation of the Monte Carlo estimator  $\hat{Y}_{MC}$  in (2.64) with  $h(X) = 1(X > a)$  is  $2.30/\sqrt{n}$ ,  $6.55/\sqrt{n}$ , and  $1870/\sqrt{n}$  for  $a = 1, 2$ , and  $5$ , respectively. Consider also an improved estimator for  $P(X > a)$  defined by

$$\hat{Y}_{IMC} = \frac{1}{n} \sum_{i=1}^n 1(X_i > a) \frac{\phi(X_i)}{\phi^*(X_i)},$$

where  $\phi(u) = \exp(-u^2/2)/\sqrt{2\pi}$ ,  $\phi^*(u) = \exp(-(u-a)^2/(2\sigma^2))/(\sqrt{2\pi}\sigma)$ ,  $\sigma > 0$ , and  $\{X_i\}$  denote independent copies of  $X$ . The coefficient of variation of  $\hat{Y}_{IMC}$  is approximately  $91.97/\sqrt{n}$ ,  $97.21/\sqrt{n}$ , and  $67.06/\sqrt{n}$  for  $a = 1, 2$ , and  $5$  if  $\sigma = 0.1$  and  $1.21/\sqrt{n}$ ,  $1.59/\sqrt{n}$ , and  $2.41/\sqrt{n}$  for  $a = 1, 2$ , and  $5$  if  $\sigma = 1.0$ . Note that  $\hat{Y}_{MC}$  deteriorates rapidly as  $a$  increases and that the efficiency of  $\hat{Y}_{IMC}$  depends strongly on the selection of sampling distribution  $\phi^*$ . ◇

Example 2.68 shows that the efficiency of improved Monte Carlo simulation depends essentially on the measure proposed for calculations. Since the coefficient of variation of  $\hat{Y}_{IMC}$  is not available analytically, the selection of an optimal density  $\phi^*(\cdot)$  requires extensive calculations. The coefficients of variation reported for  $\hat{Y}_{IMC}$  have been estimated from Monte Carlo experiments performed for various  $\sigma$ .

There is no efficient procedure for selecting measures yielding accurate and efficient estimators even for one-dimensional problems, as considered in Example 2.68. The construction of improved estimators encounters additional difficulties when dealing with multidimensional problems. For example, consider the estimation of the expectation  $E[1(X \in D)]$ , where  $X$  is an  $\mathbb{R}^d$ -valued random variable and  $D$  is a subset of  $\mathbb{R}^d$ . The selection of a new measure for  $X$  such that its samples under this measure fall in equal proportion in  $D$  and  $D^c$  is a rather complex task. The following example presents a multidimensional problem for which an improved Monte Carlo algorithm can be constructed simply.

*Example 2.69* Let  $p_s = P(X \in D)$  and  $p_f = P(X \in D^c)$ , where  $D = \{x \in \mathbb{R}^d : \|x\| \leq r\}$  is a sphere of radius  $r > 0$  centered at the origin of  $\mathbb{R}^d$  and  $X$  is an  $\mathbb{R}^d$ -valued random variable with independent  $N(0, 1)$  coordinates defined on a probability space  $(\Omega, \mathcal{F}, P)$ . The probability  $p_f$  can be calculated from

$$\begin{aligned} p_f &= \int_{\mathbb{R}^d} 1(x \in D^c) f(x) dx = E_P[1(X \in D^c)] \quad \text{or} \\ p_f &= \int_{\mathbb{R}^d} \left[ 1(x \in D^c) \frac{f(x)}{q(x)} \right] q(x) dx = E_Q \left[ 1(X \in D^c) \frac{f(X)}{q(X)} \right], \end{aligned} \quad (2.66)$$

where  $Q$  is a measure on  $(\Omega, \mathcal{F})$  such that  $P \ll Q$ . The densities  $f$  and  $q$  of the distributions induced by probability measures  $P$  and  $Q$  are

$$f(x) = (2\pi)^{-d/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^d x_i^2 \right] \quad \text{and}$$

$$q(z) = [(2\pi)\sigma^2]^{-d/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (z_1 - r)^2 + \sum_{i=2}^d z_i^2 \right] \right\}.$$

Monte Carlo estimators based on the first and the second expressions of  $p_f$  in (2.66) are denoted by  $\hat{p}_{f,MC}$  and  $\hat{p}_{f,IMC}$ , respectively. The exact probability  $p_f$  can be calculated from

$$1 - P_f = P \left( \sum_{i=1}^d X_i^2 \leq r^2 \right) = \frac{1}{\Gamma(d/2)} \int_0^{r^2/2} \xi^{d/2-1} e^{-\xi} d\xi,$$

and is equal to 0.0053,  $0.8414 \times 10^{-4}$ , and  $0.4073 \times 10^{-6}$  for  $r = 5, 6$ , and  $7$ , where  $\Gamma(\cdot)$  denotes the gamma function. Monte Carlo estimates  $\hat{p}_{f,MC}$  of  $p_f$  based on 10000 independent samples of  $X$  are 0.0053, 0.0, 0.0 for  $r = 5, 6$ , and  $7$ . Improved Monte Carlo estimates  $\hat{p}_{f,IMC}$  of  $p_f$  based on the same number of samples for  $r = 5$  are 0.0, 0.0009, 0.0053, and 0.0050 if  $\sigma = 0.5, 1.0, 2.0$ , and  $3.0$ . For  $r = 6$ , the estimates  $\hat{p}_{f,IMC}$  are  $0.0001 \times 10^{-4}$ ,  $0.1028 \times 10^{-4}$ ,  $0.5697 \times 10^{-4}$ ,  $1.1580 \times 10^{-4}$ , and  $1.1350 \times 10^{-4}$  if  $\sigma = 0.5, 1.0, 2.0, 3.0$ , and  $4.0$ . For  $r = 7$ , the estimates  $\hat{p}_{f,IMC}$  are 0.0,  $0.0016 \times 10^{-6}$ ,  $0.1223 \times 10^{-6}$ ,  $0.6035 \times 10^{-6}$ , and  $0.4042 \times 10^{-6}$  if  $\sigma = 0.5, 1.0, 2.0, 3.0$ , and  $4.0$ . Note that the density  $q$  corresponds to an  $\mathbb{R}^d$ -valued variable with independent Gaussian coordinates with variance  $\sigma^2$  and mean 0, except for a coordinate that has mean  $r$ . Monte Carlo estimates are unsatisfactory for relatively large values of  $r$ . The accuracy of the estimators  $\hat{p}_{f,IMC}$  depends essentially on  $q$ ; for example, they are unsatisfactory for  $\sigma = 0.5$  and accurate for  $\sigma$  in the range [3, 4].  $\diamond$

## 2.14 Exercises

**Exercise 2.1** Show that  $\sigma(\mathcal{A})$  defined by (2.1) is a  $\sigma$ -field, and that  $\sigma(\mathcal{A})$  is the smallest  $\sigma$ -field including  $\mathcal{A}$ .

**Exercise 2.2** Prove the properties of the probability measure  $P$  in (2.2).

**Exercise 2.3** Show that the inclusion–exclusion formula in (2.3) is valid.

*Hint:* Use the fourth formula in (2.2) to calculate the probability of  $\cup_{i=1}^m A_i$  by viewing this event as the union of  $\cup_{i=1}^{m-1} A_i$  and  $A_m$  for  $m \leq n$ .

**Exercise 2.4** Show that the conditional probability  $P(A \mid B)$  in (2.5) is a probability measure on  $(\Omega, \mathcal{F}, P)$ .

**Exercise 2.5** Prove the law of total probability and the Bayes formula in (2.6).

**Exercise 2.6** Consider the events  $A_1 = \{(6, 2)\}$ ,  $A_2 = \{(6, 2), (4, 4), (1, 6)\}$ , and  $B = \{\omega = (i, j) \in \Omega : i + j = 8\}$  in the experiment of rolling two dice. Calculate the conditional probabilities  $P(A_1 \mid B)$  and  $P(A_2 \mid B)$  by using the definition in (2.5) and by direct arguments.

**Exercise 2.7** Let  $X$  be a random element. Show that the  $\sigma(X)$  in Definition 2.13 is a  $\sigma$ -field and that this field is the smallest with respect to which  $X$  is measurable.

**Exercise 2.8** Suppose a random variable  $X$  defined on a probability space  $(\Omega, \mathcal{F}, P)$  takes a finite number of values  $a_1, \dots, a_n \in \mathbb{R}$ . Construct the  $\sigma$ -field  $\sigma(X)$  generated by this variable.

**Exercise 2.9** Prove Fatou's lemma for sequences of events, that is, show

$$P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n),$$

where  $\{A_n\}$  are events on a probability space  $(\Omega, \mathcal{F}, P)$ .

*Hint* Note that  $P(\liminf_{n \rightarrow \infty} A_n) = P(\lim_{n \rightarrow \infty} \cap_{k \geq n} A_k) = \lim_{n \rightarrow \infty} P(\cap_{k \geq n} A_k)$ , where the latter equality holds by Theorem 2.6 since  $\cap_{k \geq n} A_k$  is an increasing sequence of events. Since  $P(\cap_{k \geq n} A_k) \leq P(A_n)$ , we have  $P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n)$ . Similar arguments can be used to show  $\limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n)$ . The inequality  $\liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n)$  is valid since  $\{P(A_n)\}$  is a numerical sequence.

**Exercise 2.10** Consider two  $\mathbb{R}^d$ -valued random variables  $X$  and  $Y$  defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Show that  $P(\{X \leq x\} \cap \{Y < y\}) = P(X \leq x)P(Y < y)$  implies the independence of  $X$  and  $Y$ , where the notation  $X \leq x$  means  $(X_1 \leq x_1, \dots, X_d \leq x_d)$ .

**Exercise 2.11** Show that  $X$  defined by (2.15) is an  $\mathbb{R}^d$ -value random variable and that the collection of simple random variables constitutes a vector space.

**Exercise 2.12** Prove Jensen's inequality in (2.18) for finite-valued simple random variables.

*Hint* Use the following fact. If  $g : \mathbb{R} \rightarrow \mathbb{R}$  is convex, then  $g$  is continuous and  $g(x) = \sup\{l(x) : l(u) \leq g(u), \forall u \in \mathbb{R}\}$ , where  $l$  denotes a linear function.

**Exercise 2.13** Prove Fatou's lemma given by (2.28).

**Exercise 2.14** Show that  $X \leq Y$  a.s. and  $E[|X|], E[|Y|] < \infty$  imply  $E[X1_A] \leq E[Y1_A]$ ,  $A \in \mathcal{F}$ .

**Exercise 2.15** Show that the expectation of a positive random variable  $X$  is given by  $E[X] = \int_{[0, \infty)} P(X > x) dx$ .

*Hint* The mapping  $(x, \omega) \mapsto 1(X(\omega) > x)$  is measurable from  $([0, \infty) \times \Omega, \mathcal{B}([0, \infty)) \times \mathcal{F})$  to  $(\{0, 1\}, \mathcal{K})$ , where  $\mathcal{K} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$ . Fubini's theorem and the equality  $\int_{[0, \infty)} 1(X(\omega) > x) dx = \int_{[0, X(\omega))} dx = X(\omega)$  give

$$\begin{aligned}
\int_{[0,\infty)} P(X > x) dx &= \int_{[0,\infty)} \left[ \int_{\Omega} 1(X(\omega) > x) P(d\omega) \right] dx \\
&= \int_{\Omega} \left[ \int_{[0,\infty)} 1(X(\omega) > x) dx \right] P(d\omega) = \int_{\Omega} X(\omega) P(d\omega) = E[X].
\end{aligned}$$

**Exercise 2.16** Show that the correlation and covariance matrices of a random vector with finite variance are positive definite.

**Exercise 2.17** Prove the properties of the distribution function stated following Definition 2.32.

*Hint* Set  $B_n = \{\omega : X(\omega) \leq x_n\}$ , and  $B = \{\omega : X(\omega) \leq x\}$ , where  $\{x_n\}$  is a decreasing numerical series converging to  $x$ . The sequence of events  $B_n$  is decreasing so that  $\lim_{n \rightarrow \infty} B_n = \bigcap_{n=1}^{\infty} B_n = B$  implying  $\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(B_n) = P(\lim_{n \rightarrow \infty} B_n) = P(B) = F(x)$ .

Since  $F$  is a bounded, increasing, and right continuous function, it can only have jump discontinuities. Recall that  $F$  has a jump discontinuity at  $c$  if the left and right limits of  $F$  at  $c$  are finite but not equal. To show that  $F$  has at most a countable number of jump discontinuities, consider two distinct jump points  $\xi < \xi'$  of  $F$  and the open intervals  $I_{\xi} = (F(\xi-), F(\xi))$  and  $I_{\xi'} = (F(\xi'-), F(\xi'))$  associated with these jumps. Since  $\xi \neq \xi'$ , there exists  $\tilde{\xi} \in (\xi, \xi')$  such that  $F(\xi) \leq F(\tilde{\xi}) \leq F(\xi'-)$ , showing that  $I_{\xi}$  and  $I_{\xi'}$  are disjoint intervals. The collection of intervals  $I_{\xi}$  is countable since each  $I_{\xi}$  contains a rational number and the set of rational number is countable. The sum of all jumps of  $F$  is  $\sum_{\xi \in J} [F(\xi+) - F(\xi-)] = \sum_{\xi \in J} [F(\xi) - F(\xi-)] \leq 1$ , where  $J$  denotes the collection of jump points of  $F$ . Hence,  $\varepsilon n_{\varepsilon} \leq 1$  so that  $n_{\varepsilon} \leq 1/\varepsilon$ , where  $n_{\varepsilon}$  denotes the number of jumps of  $F$  larger than  $\varepsilon > 0$ .

**Exercise 2.18** Show that the central moments  $E[(X - \mu)^q]$  of  $X \sim N(\mu, \sigma^2)$  are zero if  $q$  is odd and equal to  $q! \sigma^q / (2^{q/2} (q/2)!)$  if  $q$  is even.

**Exercise 2.19** Let  $X \sim N(0, 1)$  and set  $Y = X^2$ . Find the covariance matrix of  $(X, Y)$ . Are  $X$  and  $Y$  correlated? Are  $X$  and  $Y$  independent?

**Exercise 2.20** Find the characteristic function of  $X = a + bN$ , where  $a, b \in \mathbb{R}$  are constants and  $N$  is a Poisson random variable with intensity  $\lambda > 0$ , that is, an  $\{0, 1, \dots\}$ -valued variable with probability  $P(N = n) = \lambda^n e^{-\lambda} / n!$ ,  $n \geq 0$ .

**Exercise 2.21** Let  $X$  and  $Y$  be random variables defined on the same probability space. Show that (1) if  $X$  and  $Y$  are independent, they are uncorrelated, (2) uncorrelated random variables can be dependent, and (3) uncorrelated Gaussian variables are independent.

**Exercise 2.22** Calculate the expectation of random variable  $X_2 \mid (X_1 = z)$  with density in (2.38).

**Exercise 2.23** Show that the characteristic function  $\varphi$  of a real-valued random variable  $X$  is positive definite.

*Hint* The function  $\varphi : \mathbb{R} \rightarrow \mathbb{C}$  is positive definite if the matrix  $\{\varphi(u_k - u_l), k, l = 1, \dots, n\}$  is positive semi-definite for all  $n \geq 1$  and  $u_k \in \mathbb{R}$ . Note that  $0 \leq E[ZZ^*] = \sum_{k,l=1}^n z_k z_l^* \varphi(u_k - u_l)$  for  $Z = \sum_{k=1}^n z_k \exp(iu_k X)$  and  $z_k \in \mathbb{C}$  arbitrary.

**Exercise 2.24** Find the expression of the characteristic function for  $X \sim N(\mu, \gamma)$  given by (2.43).

**Exercise 2.25** Calculate the moment generating function  $m(u) = E[\exp(uX)]$ ,  $u \in \mathbb{R}$ , for  $X \sim N(\mu, \sigma^2)$ .

**Exercise 2.26** Find the properties of the conditional Gaussian vector in (2.44).

**Exercise 2.27** Let  $X(\omega) = 2 + \sin(2\pi\omega)$  be a random variable defined on a probability space  $(\Omega = [0, 1], \mathcal{F} = \mathcal{B}[0, 1], P(d\omega) = d\omega)$  and let  $\Lambda_1 = [0, 1/4]$ ,  $\Lambda_2 = [1/4, 3/4]$ ,  $\Lambda_3 = [3/4, 1]$ , and  $\Lambda_4 = \{1\}$  be a measurable partition of  $\Omega$ . Calculate the conditional expectation  $E[X | \mathcal{G}]$ , where  $\mathcal{G} = \sigma(\Lambda_i, i = 1, \dots, 4)$ . Plot  $E[X | \mathcal{G}]$  and  $E[X]$  against  $\omega \in \Omega = [0, 1]$ .

**Exercise 2.28** Prove Theorem 2.15.

**Exercise 2.29** Prove the relationships (2.52) and (2.53) in Theorem 2.16.

*Hint* The defining relation gives  $\int_{\Lambda} X dP = \int_{\Lambda} E[X | \mathcal{G}_1] dP$  for all  $\Lambda \in \mathcal{G}_1$ , so that  $\int_{\Lambda} E[X | \mathcal{G}_1] dP = \int_{\Lambda} E[X | \mathcal{G}_2] dP$  under the assumption  $E[X | \mathcal{G}_1] = E[X | \mathcal{G}_2]$ . We have  $\int_{\Lambda} X dP = \int_{\Lambda} E[X | \mathcal{G}_2] dP$  for all  $\Lambda \in \mathcal{G}_1$  so that  $E[X | \mathcal{G}_2]$  is  $\mathcal{G}_1$ -measurable. Conversely, if  $E[X | \mathcal{G}_2]$  is  $\mathcal{G}_1$ -measurable, then  $E\{E[X | \mathcal{G}_2] | \mathcal{G}_1\} = E[X | \mathcal{G}_2]$ . The proof is completed by using (2.53).

Note that  $\int_{\Lambda} E\{E[X | \mathcal{G}_2] | \mathcal{G}_1\} dP = \int_{\Lambda} E[X | \mathcal{G}_2] dP = \int_{\Lambda} X dP = \int_{\Lambda} E[X | \mathcal{G}_1] dP$  holds for all  $\Lambda \in \mathcal{G}_1 \subset \mathcal{G}_2$  by the defining relation, which gives the first equality in (2.53). The second equality in this formula results since  $E[X | \mathcal{G}_1]$  is  $\mathcal{G}_1$ -measurable so that  $\mathcal{G}_2$ -measurable, which gives  $E\{E[X | \mathcal{G}_1] | \mathcal{G}_2\} = E[X | \mathcal{G}_1]$ .

**Exercise 2.30** Prove (2.55) by noting that  $\mathcal{G}$  is generated by the partition  $\{B, B^c\}$  of  $\Omega$ .

**Exercise 2.31** Show that if  $X$  and  $Z$  are random variables on the same probability space, then  $E[X | Z] = E[X | \sigma(Z)]$ , where  $E[X | Z] = \int u f_{X|Z}(u | z) du$  and  $f_{X|Z}$  is the density of  $X$  conditional on  $Z$ .

**Exercise 2.32** Let  $X$  be a real-valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$  and  $\mathcal{G}$  be a sub- $\sigma$ -field of  $\mathcal{F}$ . Show that  $E[X | \mathcal{G}] = E[X]$  for  $\mathcal{G} = \{\emptyset, \Omega\}$  and  $E[X | \mathcal{G}] = X$  for  $\mathcal{G} = \mathcal{F}$ .

*Hint* If  $\mathcal{G} = \mathcal{F}$ , then  $E[X | \mathcal{G}]$  is  $\mathcal{F}$ -measurable and  $\int_{\Lambda} (X - E[X | \mathcal{F}]) dP = 0$  for all  $\Lambda \in \mathcal{F}$  by the defining relation, so that  $X = E[X | \mathcal{F}]$  a.s.

**Exercise 2.33** Let  $X$  be a random variable with distribution  $F$  and  $a \in \mathbb{R}$  such that  $F(a) \in (0, 1)$ . Show that

$$E[X | \mathcal{G}] = \frac{\int_{-\infty}^a x dF(x)}{F(a)} 1_A + \frac{\int_a^{\infty} x dF(x)}{1 - F(a)} 1_{A^c},$$

where  $\mathcal{G} = \{\emptyset, \Omega, A, A^c\}$  and  $A = X^{-1}((-\infty, a])$ .

**Exercise 2.34** Suppose  $A_i$  and  $X_i$  in (2.57) have finite variance. Calculate the mean and variance of  $M_n$  in this equation.

**Exercise 2.35** Calculate the mean and the coefficient of variation of the estimator  $\hat{Y}_{IMC}$  in Example 2.68.

## References

1. Applebaum D (2004) Lévy processes and stochastic calculus. Cambridge University Press, Cambridge
2. Brabenec RL (1990) Introduction to real analysis. PWS-KENT Publishing Company, Boston
3. Bratley P, Fox BL, Schrage LE (1987) A guide to simulation. 2nd edn. Springer, New York
4. Chung KL (1974) A course in probability theory. Academic Press Inc., New York
5. Fristedt B, Gray L (1997) A modern approach to probability theory. Birkhäuser, Boston
6. Grigoriu M (1995) Applied non-Gaussian processes: Examples, theory, simulation, linear random vibration, and MATLAB solutions. Prentice Hall, Englewoods Cliffs
7. Grigoriu M (2002) Stochastic calculus. Applications in science and engineering. Birkhäuser, Boston
8. Hernández DB (1995) Lectures on probability and second order random fields. World Scientific, London
9. Iranpour R, Chacon P (1988) The Mark Kac lectures. Macmillan Publishing Company, New York
10. Protter P (1990) Stochastic integration and differential equations. Springer, New York
11. Resnick SI (1998) A probability path. Birkhäuser, Boston
12. Rubinstein R (1981) Simulation and the Monte Carlo method. Wiley, New York
13. Steele JM (2001) Stochastic calculus and financial applications. Springer, New York

## Chapter 3

# Random Functions

### 3.1 Introduction

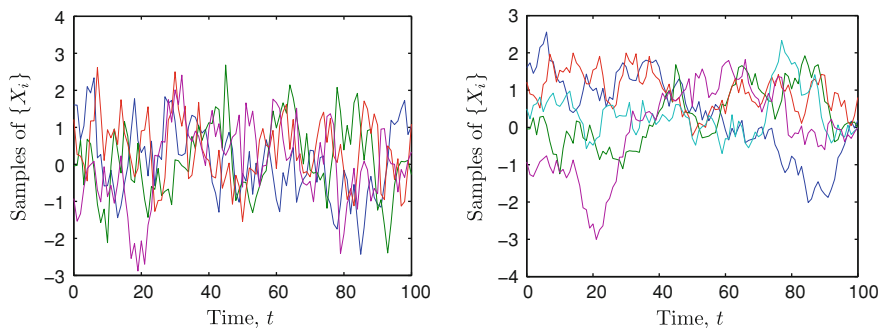
We have seen that random vectors are finite families of real- or complex-valued random variables that are completely characterized by their joint distribution. Random functions can be viewed as countable or uncountable families  $\{X_\zeta, \zeta \in I\}$  of real- or complex-valued random variables or vectors, where  $I$  is an index set.

**Definition 3.1** Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $I$  a countable or uncountable index set, and  $X : I \times \Omega \rightarrow \mathbb{R}^d$  a function of two arguments  $t \in I$  and  $\omega \in \Omega$ . We say that  $X$  is an  $\mathbb{R}^d$ -valued random function if  $X(t, \cdot)$  is an  $\mathbb{R}^d$ -valued random variable for each  $t \in I$ . If  $I$  is a countable set representing time, then  $X$  is said to be a time series. If  $I$  is an uncountable set, then  $X$  is a stochastic or random process and a random field if  $I$  is a subset of  $\mathbb{R}$  representing time and a subset of  $\mathbb{R}^{d'}$ ,  $d' \geq 1$ , representing space, respectively.

The definition extends directly to complex-valued random functions. It is common to use the abbreviated notation  $X(t)$  for random function  $X$ , rather than  $X(t, \omega)$  which indicates both arguments of  $X$ . The functions  $X(\cdot, \omega)$  indexed by  $\omega \in \Omega$  are the samples or the sample paths of  $X$ .

Recall that a random element on a probability space  $(\Omega, \mathcal{F}, P)$  is a measurable function  $X : (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S})$ , where  $S$  is a metric space and  $\mathcal{S}$  denotes the  $\sigma$ -field generated by the open sets of  $S$  (Definition 2.11). If  $S$  is a space of function depending on temporal and spatial arguments,  $X$  is a stochastic process and random field, respectively. For example,  $X(\omega)$  is a real-valued continuous function defined on  $[0, 1]$  if  $S = C[0, 1]$ . We denote by  $X(t, \omega)$  the coordinate  $t \in [0, 1]$  of  $X(\omega)$ , that is, the value of continuous function  $X(\omega)$  at  $t$ . This interpretation specifies the sample properties of  $X$  in contrast to Definition 3.1 that does not provide this information.

*Example 3.1* Let  $X(t)$ ,  $t \in [0, 1]$ , be a real-valued stochastic process with  $n < \infty$  continuous samples  $\{x_k(t)\}$ ,  $k = 1, \dots, n$ , that is defined on a probability space  $(\Omega, \mathcal{F}, P)$ , so that  $\{x_k\}$  are members of  $C[0, 1]$ . If  $X : (\Omega, \mathcal{F}) \rightarrow$



**Fig. 3.1** Samples of  $\{X_i\}$  for  $\rho = 0.7$  (left panel) and  $\rho = 0.95$  (right panel)

( $C[0, 1]$ ,  $\mathcal{C}[0, 1]$ ) is measurable, the sets  $A_k = X^{-1}(\{x_k\})$ ,  $k = 1, \dots, n$ , provide a measurable partition of  $\Omega$ , so that  $X(t, \omega) = \sum_{k=1}^n x_k(t) 1_{A_k}(\omega)$ . By analogy with random variables,  $X$  is called a simple stochastic process. Generally, a random function has an uncountable number of samples.  $\diamond$

**Definition 3.2** An  $\mathbb{R}^d$ -valued random function  $\{X(t), t \in I\}$ ,  $I \in \mathbb{R}^{d'}$ , on a probability space  $(\Omega, \mathcal{F}, P)$  is called measurable if  $X : I \times \Omega \rightarrow \mathbb{R}^d$  is a measurable from  $(I \times \Omega, \mathcal{B}(I) \times \mathcal{F})$  to  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .

If  $X$  is a measurable random function, then  $X(t, \cdot)$  is  $\mathcal{F}$ -measurable and  $X(\cdot, \omega)$  is  $\mathcal{B}(I)$ -measurable by Fubini's theorem (Theorem 2.9). The first statement is valid even if  $X$  is not measurable by the definition of random functions. There are random functions that do not have measurable versions ([1], Example 9.4.3). However, most of the random functions used in applications are measurable, for example, Brownian motion is a measurable process (Theorem 3.36).

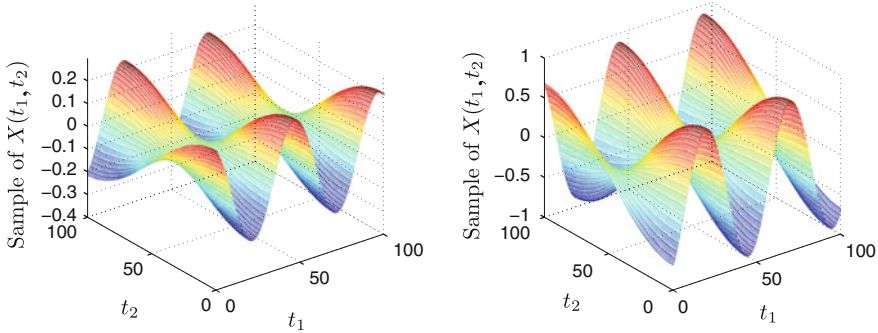
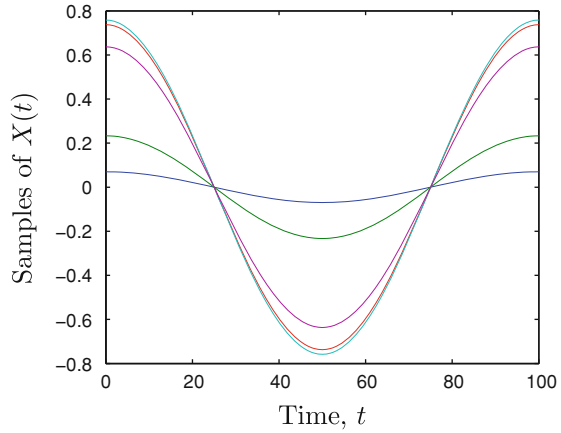
This introductory section is concluded with examples of time series, stochastic processes, and random fields. We also introduce the Brownian motion and the compound Poisson processes and show typical samples of these processes.

**Example 3.2** Let  $\{X_i\}$ ,  $i \in I = \{1, 2, \dots\}$ , be a time series defined by the recurrence formula  $X_{i+1} = \rho X_i + \sqrt{1 - \rho^2} W_i$ ,  $i = 1, 2, \dots$ , where  $|\rho| < 1$  and  $\{W_i\}$  are uncorrelated random variables with mean 0 and variance 1. If initial state  $X_1$  has mean 0 and variance 1 and is uncorrelated of  $\{W_i\}$ , then  $E[X_i X_{i+k}] = \rho^k$ . Figure 3.1 shows five samples of  $\{X_i\}$  for iid  $W_i \sim N(0, 1)$ ,  $X_1 \sim N(0, 1)$  independent of  $\{W_i\}$ ,  $\rho = 0.7$  (left panel) and  $\rho = 0.95$  (right panel). The samples of  $\{X_i\}$  for  $\rho = 0.95$  are smoother than those for  $\rho = 0.7$  since they have a stronger temporal dependence.  $\diamond$

*Proof* The expectation of  $X_{i+1} = \rho X_i + \sqrt{1 - \rho^2} W_i$  is  $E[X_{i+1}] = \rho E[X_i]$ ,  $i \in I$ , so that  $E[X_i] = 0$ ,  $i \geq 2$ , since  $E[X_1] = 0$  by assumption. The expectation of the square of this equation is  $\gamma_{i+1} = \rho^2 \gamma_i + (1 - \rho^2)$ , where  $\gamma_i = E[X_i^2]$  so that  $\gamma_{\text{st}} = \lim_{i \rightarrow \infty} \gamma_i = 1$  since  $|\rho| < 1$ . Under the assumption on



**Fig.3.2** Samples of  $X(t) = Y \cos(2\pi t)$ ,  $Y \sim U(0, 1)$



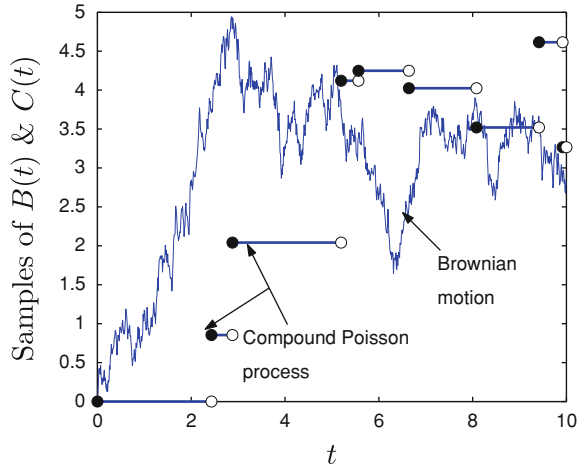
**Fig.3.3** Samples of  $X(t_1, t_2)$  for  $Y, Z \sim U(0, 1)$

$X_1$ ,  $\gamma_i = 1$  at all times. We also have  $E[X_i X_{i+1}] = E[X_i(\rho X_i + \sqrt{1 - \rho^2} W_i)] = \rho E[X_i^2]$ ,  $E[X_i X_{i+2}] = E[[X_i(\rho(\rho X_i + \sqrt{1 - \rho^2} W_i) + \sqrt{1 - \rho^2} W_{i+1})] = \rho^2 E[X_i^2]$ , and so on for all  $i \in I$  since  $X_1$  and  $\{W_i\}$  are uncorrelated by assumption.  $\blacktriangle$

*Example 3.3* The function  $X(t) = Y \cos(2\pi t)$ ,  $t \in [0, 1]$ ,  $Y \sim U(0, 1)$ , is a real-valued stochastic process since it is a random variable for each  $t \in I = [0, 1]$ . Figure 3.2 shows five independent samples of  $X$ . We say that  $X(t)$  is a parametric process since it is represented by a deterministic function of time depending on a finite number of random variables.  $\diamond$

*Example 3.4* Let  $Y, Z \sim U(0, 1)$  be independent random variables. The function  $X(t_1, t_2) = Y \cos(v_1 t_1) \cos(v_2 t_2 + 2\pi Z)$ ,  $(t_1, t_2) \in [0, 1]^2$ , is a real-valued random field, referred to parametric random field since  $X$  is a deterministic function depending on a finite number of random variables. Figure 3.3 shows two samples of this parametric random field.  $\diamond$

**Fig. 3.4** Samples of a Brownian motion and a compound Poisson process with  $Y_1 \sim N(0, \sigma^2)$  and  $\lambda\sigma^2 = 1$



**Definition 3.3** A real-valued process  $B(t)$ ,  $t \geq 0$ , is called a Brownian motion if (1)  $B(t) = 0$  a.s., that is,  $P(\{\omega \in \Omega : B(0, \omega) = 0\}) = 1$ , (2) for any  $0 \leq s < t$ ,  $B(t) - B(s) \sim N(0, t - s)$ , (3) the process has independent increments, that is, for any integer  $n \geq 1$  and  $0 \leq t_1 < t_2 < \dots < t_n$ , the random variables  $B(t_1)$ ,  $B(t_2) - B(t_1)$ ,  $\dots$ ,  $B(t_n) - B(t_{n-1})$  are independent, and (4) almost all samples of  $B$  are continuous functions, that is,  $P(\{\omega \in \Omega : B(\cdot, \omega) \text{ is continuous}\}) = 1$ .

*Example 3.5* Let  $C(t) = \sum_{k=1}^{N(t)} Y_k$ ,  $t \geq 0$ , be a compound Poisson process, where  $N$  is a homogenous Poisson counting process with intensity parameter  $\lambda > 0$  and  $\{Y_k\}$ ,  $k = 1, 2, \dots$ , are iid real-valued random variables with finite variance. Figure 3.4 shows a sample of  $C$  for  $Y_1 \sim N(0, \sigma^2)$  and  $\lambda\sigma^2 = 1$  and a sample of a Brownian process  $B(t)$ ,  $t \geq 0$ . The sample of  $C$  is piecewise constant with jumps  $\{Y_k\}$  at the jump times of  $N(t)$  while that of  $B$  is continuous.  $\diamond$

## 3.2 Finite Dimensional Distributions

Consider an  $\mathbb{R}^d$ -valued random function  $X(t)$ ,  $t \in I \subset \mathbb{R}^{d'}$ . If  $I$  is uncountable, it is not possible to specify the joint distribution of all random variables defining  $X$ . In this case,  $X$  can be characterized partially by its finite dimensional distributions.

**Definition 3.4** Let  $X(t)$ ,  $t \in I$ , be an  $\mathbb{R}^d$ -valued random function defined on a probability space  $(\Omega, \mathcal{F}, P)$ ,  $n \geq 1$  an integer, and  $t_k \in I$ ,  $k = 1, \dots, n$ , arbitrary distinct arguments. The finite dimensional distribution of order  $n$  of  $X$  is the joint distribution of the random vector  $(X(t_1), \dots, X(t_n))$ , that is,

$$F_n(x^{(1)}, \dots, x^{(n)}; t_1, \dots, t_n) = P\left(\bigcap_{k=1}^n \{X(t_k) \leq x^{(k)}\}\right) \quad (3.1)$$

where  $X(t_k) \leq x^{(k)}$  means  $X_i(t_k) \leq x_i^{(k)}$  for all  $i = 1, \dots, d$ . The distribution  $F(x; t) = F_1(x; t)$  is referred to as the marginal distribution of  $X$  at  $t$ , and represents the distribution of the  $\mathbb{R}^d$ -valued random variable  $X(t)$ .

The definition is meaningful since  $\times_{k=1}^n (-\infty, x^{(k)})$  are Borel sets so that  $\{X(t_k) \leq x^{(k)}\} \in \mathcal{F}$ . As previously,  $(-\infty, x^{(k)})$  means  $\times_{i=1}^d (-\infty, x_i^{(k)})$ . Note that  $F_n$  are probability measures induced by  $X$  on the measurable space  $(\mathbb{R}^{nd}, \mathcal{B}^{nd})$ .

**Definition 3.5** Two  $\mathbb{R}^d$ -valued random functions  $X(t)$  and  $Y(t)$ ,  $t \in I \subset \mathbb{R}^{d'}$  are said to be versions if their finite dimensional distributions coincide. Versions may or may not be defined on the same probability space.

Finite dimensional distributions provide a very useful but incomplete characterization for random functions, as illustrated by the following two examples.

*Example 3.6* Let  $X(t)$  and  $Y(t)$ ,  $0 \leq t \leq 1$ , be real-valued random functions on a probability space  $(\Omega = [0, 1], \mathcal{F} = \mathcal{B}[0, 1], P(d\omega) = d\omega)$  such that  $X(t, \omega) = 0$  and  $Y(t, \omega) = 1(t = \omega)$ . These processes have the same finite dimensional distributions but  $\sup_{t \in [0, 1]} X(t) = 0$  and  $\sup_{t \in [0, 1]} Y(t) = 1$ . Also,  $X(t)$  has continuous samples with probability 1 while almost all samples of  $Y(t)$  are not continuous, that is,  $P(X(t) \text{ continuous in } [0, 1]) = 1$  and  $P(Y(t) \text{ continuous in } [0, 1]) = 0$ .  $\diamond$

*Example 3.7* Let  $X(t)$ ,  $t \in I = [0, 1]$ , be a real-valued stochastic process with known finite dimensional distributions. This information may be insufficient for calculating statistics of some functionals of  $X$ . For example, the probability of the event  $\{|X(t)| < a, 0 \leq t \leq 1\} = \cap_{0 \leq t \leq 1} \{|X(t)| < a\}$ ,  $a > 0$ , depends on an uncountable number of random variables so that its probability cannot be obtained from the finite dimensional distributions of  $X$ . Moreover,  $\cap_{0 \leq t \leq 1} \{|X(t)| < a\}$  may not be measurable since it is the intersection of an uncountable number of events.  $\diamond$

**Definition 3.6** Let  $X$  be a real-valued random function with finite dimensional distributions in (3.1). The corresponding finite dimensional densities of  $X$  can be calculated from

$$f_n(x^{(1)}, \dots, x^{(n)}; t_1, \dots, t_n) = \frac{\partial^n F_n(x^{(1)}, \dots, x^{(n)}; t_1, \dots, t_n)}{\partial x^{(1)} \dots \partial x^{(n)}} \quad (3.2)$$

provided they exist. The marginal density of  $X$  at  $t$  is  $f(x; t) = f_1(x; t)$ . These definitions extend directly to vector- and complex-valued random functions.

Given a random function, its finite dimensional distributions of any order are well defined. However, the converse may be true, that is, it may not be possible to construct a random function whose finite dimensional distributions match an arbitrary family of probability measures.

*Example 3.8* Suppose we attempt to construct a real-valued process  $X(t)$ ,  $t \in [0, 1]$ , with finite dimensional distributions  $F_n(x_1, \dots, x_n; t_1, \dots, t_n) = \prod_{k=1}^n \Phi(x_k)$ ,  $n \geq 1$ , and continuous samples. The construction is not possible since (1) the sequence of events  $A_n = \{\omega : X(t, \omega) > \varepsilon, X(t + 1/n, \omega) < -\varepsilon\}$  converges to the

empty set as  $n \rightarrow \infty$  by the required continuity of the samples of  $X$  so that  $P(A_n) \rightarrow P(\emptyset) = 0$  as  $n \rightarrow \infty$  for every  $\varepsilon > 0$  and (2)  $P(X(t) > \varepsilon, X(s) < -\varepsilon) = (\Phi(-\varepsilon))^2$  for any  $t \neq s$  and  $\varepsilon > 0$  by the definition of  $F_n$  so that  $P(A_n) = (\Phi(-\varepsilon))^2 \not\rightarrow 0$  as  $n \rightarrow \infty$ , which leads to a contradiction.  $\diamond$

**Definition 3.7** A family of probability measures  $P_{t_1, \dots, t_n}(A_1 \times \dots \times A_n)$ ,  $A_k \in \mathcal{B}(\mathbb{R}^d)$ ,  $n = 1, 2, \dots$ ,  $d \geq 1$ , satisfies the Kolmogorov consistency criteria if

$$\begin{aligned} P_{t_1, \dots, t_n}(A_1 \times \dots \times A_n) &= P_{t_{\pi(1)}, \dots, t_{\pi(n)}}(A_{\pi(1)} \times \dots \times A_{\pi(n)}), \quad \text{and} \\ P_{t_1, \dots, t_n, t_{n+1}}(A_1 \times \dots \times A_n \times \mathbb{R}^d) &= P_{t_1, \dots, t_n}(A_1 \times \dots \times A_n), \end{aligned} \quad (3.3)$$

where  $\{\pi(1), \dots, \pi(n)\}$  is an arbitrary permutation of  $\{1, \dots, n\}$ . and  $\diamond$

If a family of probability measures satisfies the Kolmogorov consistency criteria, there exists a random function whose finite dimensional distributions match this family ([2], Sect. 1.5, [3], Theorem 1.1.16). This result is particularly useful in applications since it is common to postulate a family of probability measures, infer its unspecified parameters from the available information, and assume that the resulting measures are the finite dimensional distributions of a random function.

**Definition 3.8** A random function  $X(t)$  is said to be stationary if its finite dimensional distribution are invariant under argument translation, that is,

$$(X(t_1), \dots, X(t_n)) \stackrel{d}{=} (X(t_1 + \tau), \dots, X(t_n + \tau)), \quad (3.4)$$

or equivalently,  $F_n(x^{(1)}, \dots, x^{(n)}; t_1, \dots, t_n) = F_n(x^{(1)}, \dots, x^{(n)}; t_1 + \tau, \dots, t_n + \tau)$ , for arbitrary  $n \geq 1$ ,  $t_1, \dots, t_n \in \mathbb{R}^{d'}$ , and  $\tau \in \mathbb{R}^{d'}$ . If  $X$  is a random field satisfying (3.4), we say that  $X$  is a stationary or homogeneous random field.

*Example 3.9* Let  $(X_1, X_2, \dots)$  be a time series, where  $X_i$ ,  $i = 1, 2, \dots$ , are iid random variables. The series is stationary since the vectors  $(X_1, \dots, X_n)$  and  $(X_{1+m}, \dots, X_{n+m})$  have the same distribution for any integers  $m, n \geq 1$ .  $\diamond$

### 3.3 Sample Properties

**Definition 3.9** An  $\mathbb{R}^d$ -valued random function  $X(t)$ ,  $t \in I$ , is (1) continuous in probability at  $t \in I$  if  $\lim_{\|s-t\| \rightarrow 0} P(\|X(s) - X(t)\| \geq \varepsilon) = 0$  for all  $\varepsilon > 0$ , (2) continuous in the  $p$ 'th mean at  $t$  if  $\lim_{\|s-t\| \rightarrow 0} E[\|X(s) - X(t)\|^p] = 0$ , and (3) almost surely (a.s.) continuous at  $t$  if  $P(\{\omega : \lim_{\|s-t\| \rightarrow 0} \|X(s, \omega) - X(t, \omega)\| = 0\}) = 1$ . If  $p = 2$  in (2), we say that  $X$  is m.s. continuous at  $t$ .

A random function  $X$  is continuous in probability, continuous in the  $p$ 'th mean, and almost surely continuous in  $I$  if it is continuous in probability, continuous in the  $p$ 'th mean, and almost surely continuous at each  $t \in I$ , respectively.

Note that almost sure continuity at each time  $t$  does not imply sample continuity. If  $X$  is a.s. continuous at  $t \geq 0$ , then  $\Omega_t = \{\omega \in \Omega : \lim_{s \rightarrow t} X(s, \omega) \neq X(t, \omega)\}$  has probability zero. However, the probability of  $\cup_{t \geq 0} \Omega_t$  may not be zero or may not even be defined since this set is an uncountable union of measurable sets, which is not necessarily measurable.

*Example 3.10* Let  $(\Omega = [0, 1], \mathcal{F} = \mathcal{B}[0, 1], P(d\omega) = d\omega)$  be a probability space and let  $X$  be a real-valued stochastic process on this space defined by  $X(t, \omega) = 1(t \geq \omega)$ ,  $t \in [0, 1]$ . Since  $\Omega_t = \{t\}$ ,  $P(\Omega_t) = 0$  so that  $X$  is continuous a.s. at each  $t \in [0, 1]$ . However, the probability of  $\cup_{t \in [0, 1]} \Omega_t = [0, 1] = \Omega$  is equal to 1 so that the process is not sample continuous.  $\diamond$

*Example 3.11* Let  $C(t)$  and  $B(t)$  denote the compound Poisson and the Brownian motion processes in Example 3.5, and assume  $Y_1 \in L^2(\Omega, \mathcal{F}, P)$ . The Brownian motion and the compound Poisson processes are continuous in probability, m.s. continuous, and a.s. continuous at each  $t \geq 0$ . The Brownian motion has continuous samples while the compound Poisson process does not.  $\diamond$

*Proof* Since  $P(|B(t) - B(s)| > \varepsilon) = 2\Phi(-\varepsilon/\sqrt{|t-s|}) \rightarrow 0$  as  $|s-t| \rightarrow 0$  for any  $\varepsilon > 0$ , the Brownian motion process is continuous in probability at  $t$ . Since the event  $\{|C(t) - C(s)| > \varepsilon\}$  has non-zero probability if  $C$  has at least a jump in the time interval  $(s, t]$ , the probability that  $C$  has at least a jump in  $(s, t]$  is  $P(N(t-s) > 0) = 1 - \exp[-\lambda(t-s)]$ , and  $P(N(t-s) > 0) \rightarrow 0$  as  $|s-t| \rightarrow 0$ , the compound Poisson process is continuous in probability at  $t \geq 0$ . The processes  $B$  and  $C$  are continuous in probability at each time  $t \geq 0$ , so that they are continuous in probability in  $[0, \infty)$ .

Since the expectations  $E[(B(t) - B(s))^2] = t - s$  and  $E[(C(t) - C(s))^2] = \lambda(t-s)E[Y_1^2] + (\lambda(t-s)E[Y_1])^2$  ([4], Sect. 3.3) converge to 0 as  $|s-t| \rightarrow 0$ , the Brownian motion and the compound Poisson processes are m.s. continuous at  $t$ .

Since  $B(t) - B(t - 1/n)$  is Gaussian variable with mean zero and variance  $1/n$  for any integer  $n \geq 1$ , the events  $A_n(\varepsilon) = \{|B(t) - B(t - 1/n)| > \varepsilon\}$ ,  $n = 1, 2, \dots$ , are such that

$$P(\liminf_{n \rightarrow \infty} A_n(\varepsilon)) \leq \liminf_{n \rightarrow \infty} P(A_n(\varepsilon)) = \liminf_{n \rightarrow \infty} 2(1 - \Phi(\sqrt{n}\varepsilon)) = 0, \quad \varepsilon > 0,$$

by Fatou's lemma (Theorem 2.11), so that  $B$  is continuous a.s. at  $t$ . For  $C$ , set  $A_n = \{|C(t) - C(t - 1/n)| > 0\}$ ,  $n = 1, 2, \dots$ , and note that  $P(A_n) = 1 - e^{-\lambda/n} \rightarrow 0$  as  $n \rightarrow \infty$ . Since  $P(\liminf_n A_n) \leq \liminf_n P(A_n)$ , the probability of  $\Omega_t = \{\omega \in \Omega : \lim_{s \rightarrow t} C(s, \omega) \neq C(t, \omega)\}$  is 0, so that  $C$  is a.s. continuous at  $t$ .

The Brownian motion process has continuous samples since  $E[(B(t+h) - B(t))^4] = 3h^2$  for any  $t \geq 0$  and  $h > 0$ . So it satisfies the Kolmogorov criterium stated later in this section (Theorem 3.1) with  $\alpha = 4$ ,  $c \geq 3$ , and  $\beta = 1$ . It can be seen from Fig. 3.4 that  $C$  is not sample continuous.  $\blacktriangle$

We have seen in Examples 3.6 and 3.7 that finite dimensional distributions provide insufficient information for calculating probabilities of various functionals of random functions. Two options are available to overcome this difficulty. The first is to assume that a random function  $X(t)$ ,  $t \in I$ , satisfies certain regularity conditions so that its

samples can be characterized completely by their values on a countable set  $I^* \subset I$  that is dense in  $I$ . Under this assumption, distributions of functionals of  $X(t)$  can be obtained from values of  $X(t)$  at  $t \in I^*$ , so that they can be calculated from the finite dimensional distributions of  $X(t)$ , for example, the distribution of  $\sup_{t \in I} X(t)$  can be derived from that of  $\sup_{t \in I^*} X(t)$ .

The second option is to consider explicit analytical expressions for a random function, for example,  $X(t) = h(t, Z)$ ,  $t \in I$ , where  $Z$  is an  $\mathbb{R}^d$ -valued random variable and  $h$  is a measurable function. These representations, referred to as parametric random functions, are used extensively in applications, for example, truncated Karhunen–Loève expansions (Sect. 3.6.5) and discrete spectral representations (Sect. 3.8).

Following is a brief discussion on sample properties for real-valued random functions that includes the Kolmogorov continuity criterium. The reader is referred to [2] (Chap. 3), [5] (Sects. 3.2 and 3.3), and [1] (Chaps. 9–12) for a comprehensive discussion on sample properties of random functions.

**Definition 3.10** Let  $X(t)$  and  $Y(t)$ ,  $t \in I \subset \mathbb{R}^{d'}$ , be random functions defined on the same probability space. They are said to be indistinguishable if their samples coincide a.s., that is, the subset  $\Omega_0$  of  $\Omega$  on which the samples of  $X$  and  $Y$  differ has probability 0. The random functions  $X$  and  $Y$  are said to be modifications if  $P(\{\omega : X(t, \omega) = Y(t, \omega)\}) = 1$  at each  $t \in I$ .

In contrast to versions that may be defined on distinct probability spaces, indistinguishable processes and modifications must be defined on the same probability space. Also, indistinguishable processes are modifications but the converse is not generally true. For example, although  $\Omega_t = \{\omega : X(t, \omega) = Y(t, \omega)\}$  is measurable and  $P(\Omega_t) = 0$  at each  $t$  if  $X$  and  $Y$  are modifications, the set  $\cup_{t \in I} \Omega_t$  may not even be in  $\mathcal{F}$ .

It is common to replace a random function  $X$  by, for example, a modification of it whose samples have more desirable properties. For example, Brownian motion has a modification  $B$  that has continuous samples a.s. ([6], Theorem 26). This modification is said to be the Brownian motion process, and is used exclusively in our discussion.

**Definition 3.11** A real-valued random function  $X(t)$ ,  $t \in I \subset \mathbb{R}^{d'}$ , on a probability space  $(\Omega, \mathcal{F}, P)$  is said to be separable if there is a countable subset  $I^* \subseteq I$  and an event  $\Omega_0$  with  $P(\Omega_0) = 0$  such that for any closed subset  $B$  of  $\mathbb{R}$  the events  $\{\omega : X(t, \omega) \in B, t \in I^*\}$  and  $\{\omega : X(t, \omega) \in B, t \in I\}$  differ by a measurable subset of  $\Omega_0$ . The set  $I^*$  is called a separant for  $X$ .

We note that (1) the finite dimensional distributions of separable random functions  $X$  provide adequate information for calculating the probability of  $\sup_{t \in I} X(t)$ ,  $I \subset \mathbb{R}^{d'}$ , and other functionals of  $X(t)$  involving values of this random function over uncountable sets, (2) there are non-separable random functions ([7], p. 44), and (3) separable random functions may not have smooth samples ([1], Example 9.2.2).

**Theorem 3.1** (Kolmogorov’s continuity criterium) *Let  $X(t)$ ,  $t \in I$ , be a real-valued separable stochastic process and  $I$  a bounded interval of the real line. If there exists*

constants  $\alpha, \beta, c > 0$  such that

$$E[|X(t+h) - X(t)|^\alpha] \leq ch^{1+\beta}, \quad (3.5)$$

then  $\sup_{s,t \in I, |s-t| < h} |X(s) - X(t)| \xrightarrow{\text{a.s.}} 0$  as  $h \rightarrow 0$ , that is, every sample of  $X$  is a uniformly continuous function on  $I$  ([7], Propositions 4.1 and 4.2).

Following are facts related to separable processes that facilitate the use of Kolmogorov's continuity criterium.

**Theorem 3.2** Let  $X(t), t \in I$ , be a stochastic process defined on a probability space  $(\Omega, \mathcal{F}, P)$ . There exists a process  $\tilde{X}(t), t \in I$ , defined on the same space that is separable and  $P(X(t) = \tilde{X}(t)) = 1$  for each  $t \in I$  ([7], Proposition 2.1).

**Theorem 3.3** Let  $X(t), t \in I$ , be a separable process that is continuous in probability. Every countable set  $I^*$  dense in  $I$  is a separant ([7], Proposition 2.2).

**Theorem 3.4** Let  $X(t), t \in I$ , be a stochastic process defined on a probability space  $(\Omega, \mathcal{F}, P)$  that is continuous in probability. There exists a separable and measurable process  $\tilde{X}(t), t \in I$ , defined on the same space such that  $P(X(t) = \tilde{X}(t)) = 1$ , and any countable set  $I^*$  dense in  $I$  is a separant for  $\tilde{X}(t), t \in I$  ([7], Proposition 2.3).

*Example 3.12* Let  $X(t)$  be a separable, real-valued stationary Gaussian process with mean 0 and covariance function  $c(\tau) = E[X(t+\tau)X(t)] = \exp(-\lambda|\tau|)$ ,  $\lambda > 0$ . The process has uniformly continuous samples since the condition in (3.5) holds with  $\alpha = 4$ ,  $c = 12\lambda^2$ , and  $\beta = 1$ .  $\diamond$

*Proof* Since  $X(t+h) - X(t) \sim N(0, 2(1 - \exp(-\lambda|h|)))$  and  $1 - \exp(-|u|) \leq |u|$ , we have  $1 - \exp(-\lambda|h|) \leq \lambda|h|$  implying  $E[(X(t+h) - X(t))^4] = 12(1 - \exp(-\lambda|h|))^2 \leq 12\lambda^2|h|^2$ .  $\blacktriangle$

## 3.4 Second Moment Properties

The random functions in this section are assumed to have finite variance, so that they are members of  $L^2(\Omega, \mathcal{F}, P)$ .

**Definition 3.12** Let  $\mu(t) = E[X(t)]$ ,  $r(s, t) = E[X(s)X(t)']$ , and  $c(s, t) = E[(X(s) - \mu(s))(X(t) - \mu(t))']$  denote the mean, correlation, and covariance functions of an  $\mathbb{R}^d$ -valued random function  $X(t), t \in I \subset \mathbb{R}^{d'}$ , defined on a probability space  $(\Omega, \mathcal{F}, P)$ . The pairs  $(\mu, r)$  and  $(\mu, c)$ , referred to as the second moment properties of  $X$ , provide equivalent information since  $c(s, t) = r(s, t) - \mu(s)\mu(t)'$ .

The second moments  $E[X_i(t)^2]$  of the coordinate  $X_i(t)$  of  $X(t)$  exist and are finite since  $X \in L^2(\Omega, \mathcal{F}, P)$ . Similarly, the mean functions  $\mu_i(t) = E[X_i(t)]$  and the correlation and the covariance functions,  $r_{i,j}(s, t) = E[X_i(s)X_j(t)]$  and  $c_{i,j}(s, t) = E[(X_i(s) - \mu_i(s))(X_j(t) - \mu_j(t))]$ , exist and are finite since

$|E[X_i(s)]| \leq (E[X_i(s)^2])^{1/2}$  and  $|E[X_i(s)X_j(t)]| \leq (E[X_i(s)^2]E[X_j(t)^2])^{1/2}$  by the Cauchy–Schwarz inequality (B.17).

If  $X$  is an  $\mathbb{C}^d$ -valued random function, its second moment properties are  $\mu(t) = E[X(t)]$  and  $r(s, t) = E[X(s)(X(t)^*)']$  or  $c(s, t) = E[(X(s) - \mu(s))(X(t) - \mu(t))^*']$ , where  $z^*$  denotes the complex conjugate of  $z \in \mathbb{C}$ .

**Definition 3.13** A random function  $X$  is said to be weakly stationary if its mean is constant and its correlation function depends only on the argument difference, that is,  $\mu(t) = \mu$  is constant and  $r(s, t)$  is a function of  $s - t$  rather than  $s$  and  $t$ . Note that if  $r(s, t)$  is a function of  $s - t$  so is  $c(s, t)$  since  $\mu(t) = \mu$  is time/space invariant. We use the notations  $r(s - t)$  and  $c(s - t)$  for  $r(s, t)$  and  $c(s, t)$  if  $X$  is a weakly stationary function.

**Definition 3.14** The coordinates  $X_i$  and  $X_j$  of an  $\mathbb{R}^d$ -valued random function  $X(t)$ ,  $t \in I$ , are said to be orthogonal and uncorrelated if  $r_{i,j}(s, t) = E[X_i(s)X_j(t)] = 0$  and  $c_{i,j}(s, t) = E[(X_i(s) - \mu_i(s))(X_j(t) - \mu_j(t))] = 0$ , respectively, for all  $s, t \in I$ , where  $\mu_i(s) = E[X_i(s)]$ .

**Theorem 3.5** *The correlation function of an  $\mathbb{R}^d$ -valued random function  $X(t)$ ,  $t \in \mathbb{R}^{d'}$ , has the properties: (1)  $r_{i,j}(s, t) = r_{j,i}(t, s)$ , (2)  $|r_{i,j}(s, t)|^2 \leq r_{i,i}(s, s)r_{j,j}(t, t)$ , and (3)  $r_{i,i}(s, t)$  is positive definite. If  $X$  is weakly stationary, then  $r_{i,j}(\tau) = r_{j,i}(-\tau)$ ,  $|r_{i,j}(\tau)|^2 \leq r_{i,i}(0)r_{j,j}(0)$ , and  $r_{i,i}(\tau)$  is positive definite, where  $\tau = s - t$ . The covariance function of  $X$  has similar properties.*

*Proof* The first property follows from the definition of  $r_{i,j}$ . The second property results from the Cauchy–Schwarz inequality. The third property holds since  $0 \leq E[(\sum_{k=1}^n a_k X_i(t_k))^2] = \sum_{k,l=1}^n a_k a_l r_{i,i}(t_k, t_l)$  for any integer  $n \geq 1$ ,  $a_k \in \mathbb{R}$ , and  $t_k \in \mathbb{R}^{d'}$ . Similar arguments yield the other properties.  $\blacktriangle$

**Example 3.13** Let  $X$  be an  $\mathbb{C}^d$ -valued random function in  $L^2(\Omega, \mathcal{F}, P)$ , that is, its coordinates are complex-valued and  $|E[X_k(s)X_l(t)^*]| < \infty$  for all  $k, l = 1, \dots, d$ , and arguments  $s$  and  $t$ . The correlation functions  $r_{k,l}(t, s) = E[X_k(t)X_l(s)^*]$ ,  $k, l = 1, \dots, d$ , satisfy the condition  $r_{k,l}(t, s) = r_{l,k}(s, t)^*$ , where  $z^*$  denotes the complex conjugate of  $z \in \mathbb{C}$ .  $\diamond$

*Proof* Let  $U_k$  and  $V_k$  denote the real and imaginary parts of  $X_k$ . Then

$$\begin{aligned} r_{k,l}(s, t) &= E[(U_k(s) + iV_k(s))(U_l(t) - iV_l(t))] \\ &= E[U_k(s)U_l(t) + V_k(s)V_l(t)] + iE[V_k(s)U_l(t) - U_k(s)V_l(t)] \end{aligned}$$

and is equal to  $r_{l,k}(t, s)^* = (E[X_l(t)X_k(s)^*])^* = E[X_l(t)^*X_k(s)] = r_{k,l}(s, t)$ . Other properties of the correlation functions for complex-valued processes can be derived in a similar manner.  $\blacktriangle$

**Example 3.14** Let  $A_k, B_k$  be uncorrelated, real-valued random variables with mean zero and unit variance and let  $\sigma_k, v_k > 0$ ,  $k = 1, 2, \dots, n$ , be some constants. The function



$$X(t) = \sum_{k=1}^n \sigma_k [A_k \cos(v_k t) + B_k \sin(v_k t)], \quad t \geq 0, \quad (3.6)$$

is a real-valued parametric stochastic process with mean zero and covariance function

$$c(s, t) = \sum_{k=1}^n \sigma_k^2 \cos(v_k(s - t)), \quad (3.7)$$

so that  $X$  is weakly stationary. If in addition  $A_k, B_k$  are Gaussian variables, then  $X$  is a stationary Gaussian process (Exercise 3.2).  $\diamond$

*Proof* The definition of  $X$  and the properties of  $(A_k, B_k)$  imply  $E[X(t)] = 0$ . The covariance function of  $X$  is

$$\begin{aligned} c(s, t) &= \sum_{k,l=1}^n \sigma_k \sigma_l E[(A_k \cos(v_k s) + B_k \sin(v_k s))(A_l \cos(v_l t) + B_l \sin(v_l t))] \\ &= \sum_{k=1}^n \sigma_k^2 (\cos(v_k s) \cos(v_k t) + \sin(v_k s) \sin(v_k t)) = \sum_{k=1}^n \sigma_k^2 \cos(v_k(s - t)) \end{aligned}$$

by the linearity of the expectation operator and the properties of  $(A_k, B_k)$ .  $\blacktriangle$

*Example 3.15* An alternative representation of  $X$  in Example 3.14 is

$$X(t) = \sum_{k=-n}^n \sigma_k C_k e^{iv_k t}, \quad t \geq 0, \quad (3.8)$$

where  $C_k = (A_k - iB_k)/2$  for all  $k$  with  $A_{-k} = A_k$ ,  $B_{-k} = -B_k$ ,  $v_{-k} = -v_k$ , and  $\sigma_{-k} = \sigma_k$  for  $k = 1, \dots, n$ ,  $E[C_k] = 0$ , and  $E[C_k C_l^*] = \delta_{kl}/2$ . The correlation function of  $X$  is  $c(\tau) = E[X(t + \tau)X(t)^*] = \sum_{k=-n}^n (\sigma_k^2/2) e^{iv_k \tau}$ .  $\diamond$

*Proof* The identities  $e^{\pm i\alpha} = \cos(\alpha) \pm i \sin(\alpha)$  and (3.6) give the representation of  $X$  in (3.8).  $\blacktriangle$

*Example 3.16* The second moment properties of the Brownian motion  $B(t)$  and compound Poisson process  $C(t)$  are  $E[B(t)] = 0$ ,  $E[B(s)B(t)] = s \wedge t$ ,  $E[C(t)] = \lambda t E[Y_1]$ , and  $E[(C(s) - E[C(s)])(C(t) - E[C(t)])] = \lambda(s \wedge t) E[Y_1^2]$ , so that  $B(t)$  and  $C(t)$  are equal in the second moment sense if  $E[Y_1] = 0$  and  $\lambda E[Y_1^2] = 1$ .  $\diamond$

*Proof* Since  $B(t) \sim N(0, t)$  and has independent increments, we have  $E[B(t)] = 0$  and, for  $s < t$ ,  $E[B(s)B(t)] = E[B(s)(B(t) - B(s))] + E[B(s)^2] = E[B(s)^2] = s$ . For the compound Poisson process, we have  $E[C(t)] = E\{E[\sum_{i=1}^{N(t)} Y_i \mid N(t)]\} = E\{N(t)E[Y_1]\} = \lambda t E[Y_1]$ . For  $s < t$ ,  $E[C(s)C(t)] = E[C(s)(C(t) - C(s) + C(s))] = E[C(s)(C(t) - C(s))] + E[C(s)^2] = E[C(s)]E[C(t) - C(s)] + E[C(s)^2]$ , where the latter equality holds since the compound Poisson process has independent

increments. Also,  $E[C(s)^2] = E\{E[\sum_{i,j=1}^{N(s)} Y_i Y_j] \mid N(s)\} = E\{N(s)E[Y_1^2] + (N(s)^2 - N(s))E[Y_1]^2\} = \lambda s E[Y_1^2] + (\lambda s)^2 E[Y_1]^2$ , so that the correlation function of  $C$  is  $E[C(s)C(t)] = (\lambda s E[Y_1])(\lambda t E[Y_1]) + \lambda(s \wedge t)E[Y_1^2]$ . ▲

*Example 3.17* Let  $\zeta : \mathbb{R}^{d'} \rightarrow \mathbb{R}$  be a real-valued function that is bounded and satisfies the condition  $\int_A d\zeta(\lambda) \geq 0$  on all Borel sets in  $\mathbb{R}^{d'}$ . Then its Fourier transform

$$q(t) = \int_{\mathbb{R}^{d'}} e^{it'v} d\zeta(v), \quad t \in \mathbb{R}^{d'}, \quad (3.9)$$

is a continuous and positive definite function. ◇

*Proof* Since  $e^{it'v}$  is a continuous function of  $t$ , so is  $q$ . Note that  $q$  is positive definite if the matrix  $\{q(t_k - t_l)\}$  is Hermitian for arbitrary  $t_k \in \mathbb{R}^{d'}$ ,  $k = 1, \dots, n$ , and  $n \geq 1$ , that is,  $\sum_{k,l=1}^n a_k a_l^* q(t_k - t_l) \geq 0$ ,  $a_k \in \mathbb{C}$ . The latter sum is  $\int_{\mathbb{R}^{d'}} \sum_{k=1}^n a_k e^{it_k'v} \sum_{l=1}^n a_l^* e^{-it_l'v} d\zeta(v) = \int_{\mathbb{R}^{d'}} |\sum_{k=1}^n a_k e^{it_k'v}|^2 d\zeta(v)$ , which is positive by the postulated properties of  $\zeta$ . ▲

The converse of the statement in Example 3.17 is given by the following theorem that is particularly useful in applications involving weakly stationary random functions.

**Theorem 3.6** (Bochner's theorem) *A continuous function  $q : \mathbb{R}^{d'} \rightarrow \mathbb{C}$  is positive definite if and only if it admits the representation*

$$q(t) = \int_{\mathbb{R}^{d'}} e^{it'\lambda} d\zeta(\lambda), \quad (3.10)$$

where  $\lambda \in \mathbb{R}^{d'}$  and  $\zeta(\lambda)$  is a bounded, real-valued function satisfying  $\int_A d\zeta(\lambda) \geq 0$  for all measurable Borel sets in  $\mathbb{R}^{d'}$  ([2], Theorem 2.1.2, [8], Sect. 7.4).

## 3.5 Weakly Stationary Random Functions

This section presents properties of weakly stationary stochastic processes and random fields that are relevant for applications. The class of weakly stationary random function has been introduced in Definition 3.13.

### 3.5.1 $\mathbb{R}$ -Valued Stochastic Processes

Suppose  $X$  is a real-valued weakly stationary stochastic process with mean 0 and correlation/covariance function  $r(\tau) = c(\tau) = E[X(t + \tau)X(t)]$ . We have seen that

$r$  is positive definite. If in addition  $r$  is continuous, Bochner's theorem guarantees the existence of a real-valued, increasing, and bounded function  $S : \mathbb{R} \rightarrow \mathbb{R}$  such that

$$r(\tau) = \int_{-\infty}^{\infty} e^{i\nu\tau} dS(\nu), \quad \tau \in \mathbb{R}. \quad (3.11)$$

**Definition 3.15** The function  $S$  in (3.11) is called the spectral distribution of  $X$ . If  $S$  is absolutely continuous with respect to the Lebesgue measure on the real line, the Radon-Nikodym derivative  $s(\nu) = dS(\nu)/d\nu$  exists and is called the spectral density of  $X$ . Since  $S(\nu)$  is an increasing function,  $s(\nu) \geq 0$  and

$$r(\tau) = \int_{-\infty}^{\infty} e^{i\nu\tau} s(\nu) d\nu \quad \text{and} \quad s(\nu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\nu\tau} r(\tau) d\tau, \quad (3.12)$$

that is,  $r(\tau)$  and  $s(\nu)$  are Fourier pairs. The one-sided spectral density  $g(\nu) = 2s(\nu)1(\nu \geq 0)$  is frequently used in applications with  $\nu$  interpreted as frequency. The relationships between correlation and one-sided spectral density functions result from (3.12) and are

$$r(\tau) = \int_0^{\infty} \cos(\nu\tau) g(\nu) d\nu \quad \text{and} \quad s(\nu) = \frac{1}{\pi} \int_0^{\infty} \cos(\nu\tau) r(\tau) d\tau. \quad (3.13)$$

Note that the correlation and the spectral density functions provide equivalent information,  $s$  is an even function, that is,  $s(\nu) = s(-\nu)$ , and  $\int_{-\infty}^{\infty} s(\nu) d\nu = \int_0^{\infty} g(\nu) d\nu = E[X(t)^2] = r(0)$ .

*Example 3.18* The spectral distribution function of the weakly stationary real-valued process in Example 3.14 is  $S(\nu) = \alpha + \sum_{k=1}^n (\sigma_k^2/2)[1(\nu \geq -\nu_k) + 1(\nu \geq \nu_k)]$ , where  $\alpha$  is an arbitrary constant. This expression of  $S$  can be checked by direct calculations using (3.11). Although  $S$  is not absolutely continuous, it is common in applications to define its spectral densities by

$$s(\nu) = \frac{1}{2} \sum_{k=1}^n \sigma_k^2 [\delta(\nu - \nu_k) + \delta(\nu + \nu_k)] \quad \text{and} \quad g(\nu) = \sum_{k=1}^n \sigma_k^2 \delta(\nu - \nu_k), \quad (3.14)$$

where  $\delta(\cdot)$  is the Dirac delta function. The representations in (3.14) show that the energy of  $X$  is concentrated at a finite number of frequencies.  $\diamond$

*Example 3.19* The processes with the following correlation and spectral density functions

$$\begin{aligned} r(\tau) &= g_0 \frac{\sin(\nu_c \tau)}{\tau}, & g(\nu) &= g_0 1(0 \leq \nu \leq \nu_c) \\ r(\tau) &= \sigma^2 e^{-\lambda|\tau|}, & g(\nu) &= \frac{2\sigma^2 \lambda}{\pi(\nu^2 + \lambda^2)} \quad \text{and} \\ r(\tau) &= \sigma^2 e^{-\lambda|\tau|}(1 + \lambda|\tau|), & g(\nu) &= \frac{4\sigma^2 \lambda^3}{\pi(\nu^2 + \lambda^2)^2} \end{aligned} \quad (3.15)$$

are referred to as band limited white noise, first order Markov, and second order Markov processes. In these expressions  $g_0, v_c, \lambda, \sigma > 0$  are constants.  $\diamond$

### 3.5.2 $\mathbb{R}^d$ -Valued Stochastic Processes

Let  $X$  be an  $\mathbb{R}^d$ -valued weakly stationary process with correlation functions  $r_{k,l}(\tau) = E[X_k(t + \tau)X_l(t)]$ ,  $k, l = 1, 2, \dots, d$ . The following theorem extends the relationships in (3.12) to pairs of distinct coordinates of  $X$ .

**Theorem 3.7** *If  $X$  is an  $\mathbb{R}^d$ -valued stochastic process with mean 0 and continuous correlation function  $r(\tau) = E[X(t + \tau)X(t)']$ , then*

$$r_{k,l}(\tau) = \int_{\mathbb{R}} e^{iv\tau} dS_{k,l}(v) = \int_{\mathbb{R}} e^{iv\tau} s_{k,l}(v) dv, \quad \text{where} \\ S_{k,l} = S_{l,k}^* = \frac{1}{2} [S_1 - iS_2 - (1-i)(S_{k,k} + S_{l,l})], \quad k \neq l, \quad (3.16)$$

$S_{k,k}$  is the spectral distribution function of  $X_k$ ,  $s_{k,l}(v) = dS_{k,l}(v)/dv$  provided it exists, and  $S_p$  denotes the spectral distribution of  $Y_p(t) = \sum_{k=1}^d \zeta_{p,r} X_r(t)$ ,  $p = 1, 2$ , where  $\zeta_{1,k} = \zeta_{1,l} = 1$  and  $\zeta_{1,q} = 0$ ,  $q \neq k, l$ , and  $\zeta_{2,k} = i$ ,  $\zeta_{2,l} = 1$ , and  $\zeta_{2,q} = 0$ ,  $q \neq k, l$ , for an arbitrary pair  $(k, l)$  of coordinates of  $X$ .

*Proof* The representation of  $r_{k,k}$  follows from (3.12). The correlation functions of  $Y_1(t)$  and  $Y_2(t)$  are

$$r_{y_1}(\tau) = r_{k,k}(\tau) + r_{l,l}(\tau) + r_{k,l}(\tau) + r_{l,k}(\tau) \quad \text{and} \\ r_{y_2}(\tau) = r_{k,k}(\tau) + r_{l,l}(\tau) + ir_{k,l}(\tau) - ir_{l,k}(\tau),$$

and have the representations  $r_{y_p}(\tau) = \int_{-\infty}^{\infty} e^{iv\tau} dS_p(v)$ ,  $p = 1, 2$ , which together with the previous equations and Bochner's theorem give

$$r_{k,l}(\tau) = \int_{\mathbb{R}} e^{iv\tau} dS_{k,l}(v), \quad k \neq l, \quad \text{where} \\ S_{k,l} = S_{l,k}^* = \frac{1}{2} [S_1 - iS_2 - (1-i)(S_{k,k} + S_{l,l})].$$

The spectral distributions  $S_{k,l}$ ,  $k \neq l$ , are bounded complex-valued functions and the matrix  $\{S_{k,l}\}$  is Hermitian since  $S_{k,l} = S_{l,k}^*$ .

It remains to show that it is possible to define a spectral density for distinct pairs of coordinates of  $X$ . Consider a bounded interval  $[v, v + \Delta v]$  and set  $\Delta S_{k,l}(v) = S_{k,l}(v + \Delta v) - S_{k,l}(v)$ . The Hermitian matrix  $\{\Delta S_{k,l}(v)\}$ ,  $k, l = 1, \dots, d$ , has the property  $\sum_{k,l=1}^d z_k z_l^* \Delta S_{k,l}(v) \geq 0$ ,  $z_k \in \mathbb{C}$ , since the correlation function of  $Y(t) = \sum_{k=1}^d z_k X_k(t)$ ,  $z_k \in \mathbb{C}$ , is

$$r_y(\tau) = \int_{-\infty}^{\infty} e^{iv\tau} dS_y(v) = \sum_{k,l=1}^n z_k z_l^* \int_{-\infty}^{\infty} e^{iv\tau} dS_{k,l}(v),$$

and  $S_y$  is an increasing function. For distinct indices  $k, l \in \{1, \dots, d\}$  and  $z_q = 0$  for  $q \neq k, l$  the sum  $\sum_{k,l=1}^d z_k z_l^* \Delta S_{k,l}(v)$  becomes

$$\begin{aligned} \Delta S_{k,k}(v)|z_k|^2 + \Delta S_{l,l}(v)|z_l|^2 + 2\Re(\Delta S_{k,l}(v)z_k z_l^*) &\geq 0, \quad \text{or} \\ \Delta S_{k,k}(v)|z_k|^2 + \Delta S_{l,l}(v)|z_l|^2 + 2|\Delta S_{l,l}(v)||z_k||z_l^*| &\geq 0, \end{aligned}$$

because the real part of  $\Delta S_{k,l}(v)z_k z_l^*$  is smaller than its absolute value. The latter expression divided by  $|z_l|^2$  is a polynomial of  $\eta = |z_k|/|z_l|$  that has no real roots so that  $|\Delta S_{k,l}(v)|^2 \leq |\Delta S_{k,k}(v)||\Delta S_{l,l}(v)|$ . Hence,  $s_{k,l}(v)$  exists and satisfies the condition  $|s_{k,l}(v)|^2 \leq s_{k,k}(v)s_{l,l}(v)$ ,  $v \in \mathbb{R}$  ([8], Sect. 8.1).  $\blacktriangle$

*Example 3.20* Let  $X_k(t) = \sqrt{1-\rho}W_k(t) + \sqrt{\rho}W(t)$ ,  $k = 1, 2$ , be the coordinates of an  $\mathbb{R}^2$ -valued process  $X$ , where  $\rho \in (0, 1)$  and  $(W_1, W_2, W)$  are zero-mean, weakly stationary, mutually uncorrelated processes. Then  $X$  is a weakly stationary process with spectral densities  $s_{k,k}(v) = (1-\rho)s_{W_k}(v) + \rho s_W(v)$ ,  $k = 1, 2$ , and  $s_{1,2}(v) = s_{2,1}(v) = \rho s_W(v)$ , where  $s_{W_k}$  and  $s_W$  denote the spectral densities of  $W_k$  and  $W$ , respectively.  $\diamond$

*Proof* The correlation functions of  $X$  are  $r_{k,l}(\tau) = (1-\rho)\delta_{kl}r_{W_k}(\tau) + \rho r_W(\tau)$ , where  $r_{W_k}$  and  $r_W$  denote the correlation functions of  $W_k$  and  $W$ . The spectral densities  $s_{k,k}$  are the Fourier transform of the correlation functions  $r_{k,k}$  by Bochner's theorem. The correlation functions of processes  $Y_p$  with spectral distributions  $S_p$ ,  $p = 1, 2$ , in (3.16) are  $r_{y_1}(\tau) = (1-\rho)(c_{W_1}(\tau) + c_{W_2}(\tau)) + 4\rho c_W(\tau)$  and  $r_{y_2}(\tau) = (1-\rho)(c_{W_1}(\tau) + c_{W_2}(\tau)) + 2\rho c_W(\tau)$ , so that

$$\begin{aligned} s_1(v) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iv\tau} r_{y_1}(\tau) d\tau = (1-\rho)(s_{W_1}(v) + s_{W_2}(v)) + 4\rho s_W(v), \\ s_2(v) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iv\tau} r_{y_2}(\tau) d\tau = (1-\rho)(s_{W_1}(v) + s_{W_2}(v)) + 2\rho s_W(v), \quad \text{and} \end{aligned}$$

$s_{1,2}$  results from (3.16).  $\blacktriangle$

### 3.5.3 $\mathbb{R}$ -Valued Random Fields

Consider a real-valued random field  $X$  defined on  $\mathbb{R}^{d'}$  with mean, correlation, and covariance functions  $\mu(t) = E[X(t)]$ ,  $r(s, t) = E[X(s)X(t)]$ , and  $c(s, t) = E[(X(s) - \mu(s))(X(t) - \mu(t))]$ . If  $X$  is weakly stationary/homogeneous, then  $\mu$  is constant and  $r$  and  $c$  depend on  $s - t$ .

If  $X$  is weakly homogeneous and its correlation function is continuous, Bochner's theorem applies and gives

$$r(\tau) = \int_{\mathbb{R}^{d'}} e^{i\tau'v} dS(v) = \int_{\mathbb{R}^{d'}} e^{i\tau'v} s(v) dv, \quad (3.17)$$

where  $S$  denotes the spectral distribution of  $X$  ([2], Theorem 2.1.2). If  $S$  is absolutely continuous, then  $X$  has the spectral density  $s(v) = \partial^{d'} S(v) / \partial v_1 \cdots \partial v_{d'}$ . If  $X$  has a spectral density, then its correlation and spectral density functions are Fourier pairs, that is,

$$r(\tau) = \int_{\mathbb{R}^{d'}} e^{i\tau'v} s(v) dv \quad \text{and} \quad s(v) = \frac{1}{(2\pi)^{d'}} \int_{\mathbb{R}^{d'}} e^{-i\tau'v} r(\tau) d\tau. \quad (3.18)$$

**Theorem 3.8** *Let  $X(t)$ ,  $t \in \mathbb{R}^{d'}$ , be a real-valued homogeneous random field with correlation and spectral density function in (3.18). Then  $r(\tau) = r(-\tau)$  and  $s(v) = s(-v)$  for all  $\tau$  and  $v$ .*

*Proof* We have  $r(\tau) = E[X(t+\tau)X(t)] = E[X(t)X(t-\tau)] = E[X(t-\tau)X(t)] = r(-\tau)$  since  $X$  is homogeneous. Also,

$$\begin{aligned} s(-v) &= \frac{1}{(2\pi)^{d'}} \int_{\mathbb{R}^{d'}} e^{i\tau'v} r(\tau) d\tau \\ &= \frac{1}{(2\pi)^{d'}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-i\tau'v} r(-\tau) (-d\tau_1) \cdots (-d\tau_{d'}) = s(v), \end{aligned}$$

by the change of variable  $\tau \mapsto -\tau$  and the property  $r(\tau) = r(-\tau)$ .  $\blacktriangle$

*Example 3.21* Let  $X$  be a real-valued, weakly stationary random field defined on  $\mathbb{R}^2$  with mean 0, variance 1, and correlation function

$$r(\tau) = E[X(t+\tau)X(t)] = \exp \left[ -\frac{1}{2}(\tau_1^2 + 2\rho\tau_1\tau_2 + \tau_2^2) \right], \quad \tau \in \mathbb{R}^2, \quad (3.19)$$

where  $|\rho| < 1$ . Since  $r$  is continuous, it admits the representation in Bochner's theorem with spectral density

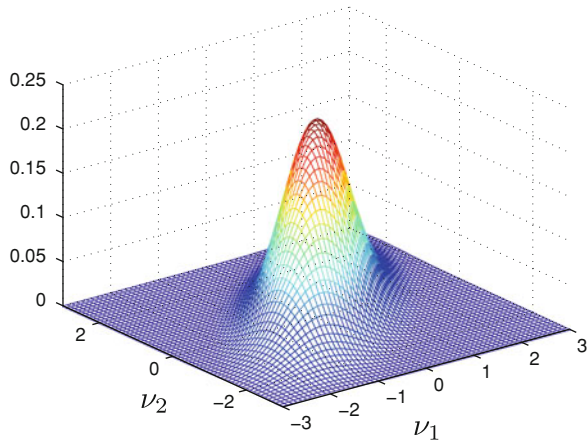
$$s(v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left[ -\frac{v_1^2 - 2\rho v_1 v_2 + v_2^2}{2(1-\rho^2)} \right], \quad v \in \mathbb{R}^2, \quad (3.20)$$

that is shown in Fig. 3.5 for  $\rho = 0.7$ .  $\diamond$

*Example 3.22* Let  $X_k$  be real-valued weakly stationary random fields with mean 0 and correlation functions  $r_k(\tau_k) = E[X_k(t_k)X_k(t_k + \tau_k)]$ , where  $t_k, \tau_k \in \mathbb{R}$  and  $k = 1, \dots, d'$ . If  $\{X_k\}$  are mutually independent, then  $X(t) = \prod_{k=1}^{d'} X_k(t_k)$  is a real-valued weakly stationary random field with mean 0 and correlation function  $r(\tau) = \prod_{k=1}^{d'} r_k(p_k(\tau))$ , where  $t = (t_1, \dots, t_{d'})$  and  $p_k(\tau) = \tau_k$ ,  $k = 1, \dots, d'$ , are projection functions and  $\tau = (\tau_1, \dots, \tau_{d'}) \in \mathbb{R}^{d'}$ .  $\diamond$

Arguments similar to those used for  $\mathbb{R}^d$ -valued stochastic processes can be employed to define correlation/covariance functions and spectral densities for vector-valued random fields. Representations as in (3.16) can be constructed for the correlation functions of the same and distinct coordinates of  $\mathbb{R}^d$ - or  $\mathbb{C}^d$ -valued random

**Fig. 3.5** Spectral density  $s(v)$  of  $X$  with  $\rho = 0.7$



fields. The remainder of this section considers a special class of weakly homogeneous random fields, referred to as isotropic random fields.

**Definition 3.16** A real-valued random field  $X$  defined in  $\mathbb{R}^{d'}$  is said to be weakly isotropic if its correlation function  $r(s, t) = E[X(s)X(t)] = r(\|s - t\|)$  depends only on the length  $\|s - t\|$  of vector  $s - t$ .

The correlation function of these fields is subjected to various constraints [9], of which some are stated by the following theorem. Note that a weakly isotropic random field is homogeneous, but the converse is not true.

**Theorem 3.9** The correlation function  $r(\tau)$  of a real-valued, weakly isotropic random field  $X(t)$ ,  $t \in \mathbb{R}^{d'}$ , with mean 0 is such that  $r(\tau) \geq -r(0)/d'$ .

*Proof* Let  $t_1, \dots, t_{d'+1}$  be  $d' + 1$  points in  $\mathbb{R}^{d'}$  such that  $\|t_i - t_j\| = a > 0$  for all  $i \neq j$ . The stated inequality follows from  $E[|\sum_{k=1}^{d'+1} X(t_k)|^2] = (d' + 1)[r(0) + d'r(a)]$  and  $E[|\sum_{k=1}^{d'+1} X(t_k)|^2] \geq 0$ . ▲

**Theorem 3.10** Let  $X$  be a real-valued, weakly isotropic random field defined on  $\mathbb{R}^{d'}$  with continuous correlation function admitting a spectral density  $s(v)$ . Then the correlation and spectral density of  $X$  satisfy relationships as in (3.18), and  $s(v)$  is a function of  $\|v\|$ .

*Proof* Since  $X$  is an isotropic field,  $r(t) = r(\varphi(t))$  for all  $t \in \mathbb{R}^{d'}$ , where  $\varphi(t) = (\sum_{j=1}^{d'} a_{1j}t_j, \dots, \sum_{j=1}^{d'} a_{d'j}t_j)' = at$  defines a rotation, so that  $a = \{a_{ij}\}$  is such that  $a' = a^{-1}$  and  $a'a = aa' = I$  is the identity matrix. Since  $r(t) = \int_{\mathbb{R}^{d'}} \exp(it'v) dS(v) = \int_{\mathbb{R}^{d'}} \exp(i\varphi(t)'v) dS(v) = r(\varphi(t))$  by assumption, and  $\varphi(t)'v = (at)'v = t'a'v = t'\psi(v)$ , we have  $\int_{\mathbb{R}^{d'}} \exp(it'\psi(v)) dS(v) = \int_{\mathbb{R}^{d'}} \exp(it'\lambda) dS(\psi^{-1}(\lambda))$ , where  $v = \psi^{-1}(\lambda)$ . This shows that the spectral density is invariant to rotation so that it depends only on  $\|v\|$  ([2], Sect. 2.5). ▲

### 3.6 Second Moment Calculus

The Chebyshev and Cauchy–Schwarz inequalities and the following theorem giving properties of mean square convergent sequences are the main tools for second moment calculus.

**Theorem 3.11** *If  $X$ ,  $\{X_n, n \geq 1\}$ ,  $Y$ , and  $\{Y_n, n \geq 1\}$  are random variables with finite variance such that  $X_n \xrightarrow{\text{m.s.}} X$  and  $Y_n \xrightarrow{\text{m.s.}} Y$ , then*

$$\begin{aligned} (1) \quad & \lim_{n \rightarrow \infty} E[X_n] = E[\text{l.i.m.}_{n \rightarrow \infty} X_n] = E[X], \\ (2) \quad & \lim_{m, n \rightarrow \infty} E[X_m Y_n] = E[(\text{l.i.m.}_{m \rightarrow \infty} X_m)(\text{l.i.m.}_{n \rightarrow \infty} Y_n)] = E[XY], \text{ and} \\ (3) \quad & \text{If } X_n \xrightarrow{\text{m.s.}} X \text{ and } X_n \xrightarrow{\text{m.s.}} Z, \text{ then } P(X \neq Z) = 0, \end{aligned} \tag{3.21}$$

where  $\text{l.i.m.}_{n \rightarrow \infty} X_n = X$  means  $E[(X_n - X)^2] \rightarrow 0, n \rightarrow \infty$ .

*Proof* Since  $0 \leq |E[X] - E[X_n]| = |E[X - X_n]| \leq (E[(X - X_n)^2])^{1/2}$  holds by properties of expectation and the Cauchy–Schwarz inequality and  $X_n \xrightarrow{\text{m.s.}} X$  by assumption, we have the convergence  $E[X_n] \rightarrow E[X]$  stated in (1). For (2), note that

$$\begin{aligned} |E[X_m Y_n - XY]| &= |E[X_m Y_n - X_m Y + X_m Y - XY]| \\ &\leq |E[X_m(Y_n - Y)]| + |E[(X_m - X)Y]| \\ &\leq \left(E[X_m^2]E[(Y_n - Y)^2]\right)^{1/2} + \left(E[(X_m - X)^2]E[Y^2]\right)^{1/2}, \end{aligned}$$

by properties of expectation and the Cauchy–Schwarz inequality. The postulated m.s. convergence implies  $|E[X_m Y_n - XY]| = |E[X_m Y_n] - E[XY]| \rightarrow 0$  as  $m, n \rightarrow \infty$ . For (3), the inequalities  $0 \leq E[(X - Z)^2] = E[((X - X_n) + (X_n - Z))^2] \leq 2E[(X - X_n)^2] + 2E[(X_n - Z)^2]$  following from  $(a + b)^2 \leq 2a^2 + 2b^2$  and the Chebyshev inequality  $P(|X - Z| > \varepsilon) \leq E[(X - Z)^2]/\varepsilon^2, \varepsilon > 0$ , imply  $P(|X - Y| > \varepsilon) = 0, \forall \varepsilon > 0$ , that is, the uniqueness with probability 1 of m.s. limit.  $\blacktriangle$

The theorem shows that expectation and m.s. limit can be interchanged under some conditions and that m.s. limit is unique with probability one, that is, the measure of the subset of  $\Omega$  in which m.s. limits of a sequence  $X_n$  may differ is zero. The following section defines m.s. continuity, differentiation, and integration, give criteria under which these operations can be performed, and present examples.

#### 3.6.1 Continuity

**Definition 3.17** A real-valued random function  $X(t), t \in \mathbb{R}^{d'}$ , with finite variance is m.s. continuous or continuous in the mean square sense at  $t \in \mathbb{R}^{d'}$  if



*l.i.m.* $_{\|s-t\| \rightarrow 0} X(s) = X(t)$ , that is,  $\lim_{\|s-t\| \rightarrow 0} E[(X(s) - X(t))^2] = 0$ , where  $\|\cdot\|$  denotes a norm in  $\mathbb{R}^{d'}$ . The random function  $X$  is m.s. continuous in  $I$  if it is m.s. continuous at every point in  $I \subset \mathbb{R}^{d'}$ . An  $\mathbb{R}^d$ -valued random function  $X$  is m.s. continuous if its coordinates are m.s. continuous.

**Theorem 3.12** *A real-valued random function  $X(t)$ ,  $t \in I \subset \mathbb{R}^{d'}$ , is m.s. continuous at  $t \in I$  if and only if its correlation function  $r(u, v) = E[X(u)X(v)]$  is continuous at  $u = v = t$ , that is,  $\lim_{\|u-t\|, \|v-t\| \rightarrow 0} r(u, v) = r(t, t)$ .*

*Proof* We have

$$\begin{aligned} |r(u, v) - r(t, t)| &= |E[X(u)X(v)] - E[X(t)X(u)] + E[X(t)X(u)] - E[X(t)X(t)]| \\ &\leq |E[X(u)(X(v) - X(t))]| + |E[X(t)(X(u) - X(t))]| \\ &\leq (E[X(u)^2]E[(X(v) - X(t))^2])^{1/2} + (E[X(t)^2]E[(X(u) - X(t))^2])^{1/2} \rightarrow 0 \end{aligned}$$

as  $\|u - t\|, \|v - t\| \rightarrow 0$  since  $X$  is m.s. continuous at  $t \in I$  by assumption.

Conversely, if  $r(u, v) \rightarrow r(t, t)$  as  $\|u - t\|, \|v - t\| \rightarrow 0$ , the mean square difference  $E[(X(s) - X(t))^2] = r(s, s) + r(t, t) - 2r(s, t) \rightarrow 0$  implies the m.s. convergence of  $X$  at  $t \in I$ .  $\blacktriangle$

*Example 3.23* A real-valued weakly stationary random function  $X$  is m.s. continuous at  $t$  if and only if  $\lim_{\tau \rightarrow 0} r(\tau) = r(0)$ ,  $\tau \in \mathbb{R}^{d'}$  (Theorem 3.12). For example, a stochastic process  $X$  with correlation function  $r(\tau) = \exp(-\lambda|\tau|)$ ,  $\lambda > 0$ , is m.s. continuous over the entire real line since  $r(\tau)$  is continuous at 0.

Similarly, the weakly stationary stochastic process  $X(t)$  in Example 3.14 is m.s. continuous at any  $t \in \mathbb{R}$  since its covariance function  $c(\tau) = \sum_{k=1}^n \sigma_k^2 \cos(v_k \tau)$  is continuous at  $\tau = 0$ .  $\diamond$

*Example 3.24* Let  $X(t) = \sum_{i=1}^{N(t)} Y_i$ ,  $t \geq 0$ , where  $N(t)$  is a homogeneous Poisson process with intensity  $\lambda > 0$  and  $\{Y_i\}$  are iid random variables with finite variance. The mean and covariance functions of  $X(t)$  are  $E[X(t)] = \lambda t E[Y_1]$  and  $c(s, t) = \lambda(s \wedge t) E[Y_1^2]$  ([4], Sect. 3.3). Since  $c(u, v) \rightarrow c(t, t)$  as  $u, v \rightarrow t$  implying  $r(u, v) \rightarrow r(t, t)$  as  $u, v \rightarrow t$ ,  $X$  is m.s. continuous in  $[0, \infty)$ . However, the samples of  $X$  have discontinuities at the jump times of  $N$  showing that m.s. continuity provides limited if any information on sample properties.  $\diamond$

### 3.6.2 Differentiability

**Definition 3.18** Let  $X(t)$ ,  $t \in \mathbb{R}^{d'}$ , be a real-valued function with finite variance and  $\delta_i \in \mathbb{R}^{d'}$  a vector with coordinate  $i \in \{1, \dots, d'\}$  equal to 1 and all other coordinates equal to 0. The m.s. derivative of  $X$  with respect to coordinate  $i$  at  $t$ , denoted by  $\dot{X}_i(t)$ , is the m.s. limit of  $(X(t + h\delta_i) - X(t))/h$  as  $h \rightarrow 0$  provided it exists. If  $d' = 1$ , we denote  $\dot{X}_i$  by  $\dot{X}$ , for example, the case of real-valued stochastic processes.

It is not possible to prove the existence of m.s. derivative by calculating the expectation of  $(\dot{X}_i(t) - (X(t + h\delta_i) - X(t))/h)^2$  since the limit  $\dot{X}_i(t)$  is not known. We need to show that  $\{(X(t + h_n\delta_i) - X(t))/h_n\}$  is a Cauchy sequence in  $L^2$ , so that it has a unique limit in this space as  $h_n \downarrow 0$ ,  $n \rightarrow \infty$ , and this limit defines  $\dot{X}_i$ . As for continuity, an  $\mathbb{R}^d$ -valued random function is m.s. differentiable if its coordinates have this property.

**Theorem 3.13** *If  $\partial^2 r(u, v)/\partial u_i \partial v_i$  exists and is finite at  $(t, t) \in \mathbb{R}^{2d'}$ , then  $\dot{X}_i(t)$ ,  $i = 1, \dots, d'$  exists. The correlation function of  $\dot{X}_i$  is  $E[\dot{X}_i(s)\dot{X}_i(t)] = \partial^2 r(s, t)/\partial s_i \partial t_i$  ([2], Theorem 2.2.2 and [10], Sect. 4.4). If  $X$  is weakly stationary, the requirements on  $\partial^2 r(u, v)/\partial u_i \partial v_i$  need to be satisfied only at the origin since  $r(u, v)$  depends only on  $u - v$ .*

*Proof* Direct calculations give  $E[Y_i(s)Y_i(t)] \rightarrow \partial^2 r(s, t)/\partial s_i \partial t_i$  as  $h \rightarrow 0$ , where  $Y_i(s) = (X(s + h\delta_i) - X(s))/h$  and  $h > 0$ .  $\blacktriangle$

*Example 3.25* The weakly stationary process  $X$  in Example 3.23 with exponential correlation is m.s. continuous. However, it is not m.s. differentiable since its correlation function is not twice differentiable at the origin.  $\diamond$

*Example 3.26* Let  $X$  be a real-valued m.s. differentiable stochastic process. Then, differentiation and expectation operators commute, that is,

$$\begin{aligned} \frac{d}{dt}E[X(t)] &= E[\dot{X}(t)], \quad \frac{\partial}{\partial t}E[X(t)X(s)] = E[\dot{X}(t)X(s)], \quad \text{and} \\ \frac{\partial^2}{\partial t \partial s}E[X(t)X(s)] &= E[\dot{X}(t)\dot{X}(s)] \quad \text{or} \quad r_{\dot{X}, \dot{X}}(t, s) = \frac{\partial^2 r(t, s)}{\partial t \partial s}. \end{aligned} \quad (3.22)$$

These results extend directly to calculate expectations of higher order derivatives for real-valued processes and random fields.  $\diamond$

*Proof* Since  $\dot{X}$  exists, the sequence  $Y_n = (X(t + h_n) - X(t))/h_n$  converges in mean square to  $\dot{X}(t)$  as  $n \rightarrow \infty$  implying  $\lim_{n \rightarrow \infty} E[Y_n] = E[\text{l.i.m.}_{n \rightarrow \infty} Y_n] = E[\dot{X}(t)]$ , where  $h_n \downarrow 0$  as  $n \rightarrow \infty$  (Theorem 3.11). This fact and the equality  $E[Y_n] = (E[X(t + h_n)] - E[X(t)])/h_n$  prove the first equality in (3.22). Similar arguments can be used to obtain the other formulas in this equation.  $\blacktriangle$

*Example 3.27* If  $X$  is a m.s. differentiable, real-valued, weakly stationary process, the correlation function of  $\dot{X}(t)$  can be calculated from  $r_{\dot{X}, \dot{X}}(\tau) = -r''(\tau)$  or  $r_{\dot{X}, \dot{X}}(\tau) = -\int_{-\infty}^{\infty} v^2 e^{iv\tau} dS(v)$ , where  $S(v)$  and  $r(\tau)$  denote the spectral distribution and the correlation function of  $X$ . These results follow from the previous example and Bochner's theorem. If  $r(\tau) = (1 + \lambda|\tau|) \exp(-\lambda|\tau|)$ ,  $\lambda > 0$ , we have  $r''(\tau) = -\lambda^2(1 - \lambda|\tau|) \exp(-\lambda|\tau|)$  and  $E[\dot{X}(t)^2] = -r''(0) = \lambda^2$ , so that  $X$  with this covariance function is m.s. differentiable.  $\diamond$

*Example 3.28* The correlation function of a standard Brownian motion  $B(t)$  and a compound Poisson process  $C(t)$  with  $E[Y_1] = 0$  and  $\lambda E[Y_1^2] = 1$  is  $r(u, v) = E[B(u)B(v)] = E[C(u)C(v)] = u \wedge v$ , so that these processes are not m.s. differentiable. Formal calculations give  $\partial^2 r(u, v)/(\partial u \partial v) = \delta(u - v)$ , where  $\delta(\cdot)$  denotes

the Dirac delta function, so that the one-sided spectral density of  $\dot{B}(t)$  and  $\dot{C}(t)$  is  $g(v) = 1/\pi$ ,  $v \geq 0$ . Processes with constant spectral density and delta correlation functions are called white noise processes in the engineering and physics literature. The above calculations are meaningless since  $\dot{B}(t)$  and  $\dot{C}(t)$  do not exist in the m.s. sense. We will consider alternative ways of defining white noise processes based on properties of the Brownian, compound Poisson, and other processes.  $\diamond$

### 3.6.3 Integration

We examine briefly integrals whose integrands or integrators are real-valued stochastic processes with finite variance, that is, integrals of the type  $\int_a^b h(t) dX(t)$  and  $\int_a^b X(t) dh(t)$ , where  $h : [a, b] \rightarrow \mathbb{R}$  is a real-valued deterministic function and  $X$  denotes a real-valued stochastic process. Similar integrals can be defined for random fields ([2], Sect. 2.3).

**Definition 3.19** Let  $p_n = (a = t_0 < t_1 < \cdots < t_n = b)$  be a sequence of partitions of  $[a, b]$  with intermediate points  $t'_k \in [t_{k-1}, t_k]$  and mesh  $\Delta(p_n) = \max_{1 \leq k \leq n} (t_k - t_{k-1}) \rightarrow 0$  as  $n \rightarrow \infty$ . If

$$\begin{aligned} S_{h,X}(p_n) &= \sum_{k=1}^n h(t'_k) (X(t_k) - X(t_{k-1})) \quad \text{and} \\ S_{X,h}(p_n) &= \sum_{k=1}^n X(t'_k) (h(t_k) - h(t_{k-1})). \end{aligned} \quad (3.23)$$

are Cauchy sequences in  $L^2$ , their m.s. limits are denoted by  $\int_a^b h(t) dX(t)$  and  $\int_a^b X(t) dh(t)$ , and are called m.s. Riemann–Stieltjes integrals of  $h$  with respect to  $X$  and  $X$  with respect to  $h$ , respectively.

Most conditions for the existence of the integrals  $\int_a^b h(t) dX(t)$  and  $\int_a^b X(t) dh(t)$  involve metrics quantifying the variation of  $h$  and  $X$  over a partition  $p_n = (a = t_0 < t_1 < \cdots < t_n = b)$  of  $[a, b]$  ([5], Sect. 3.9.3.1).

**Definition 3.20** The variations of  $h$  and  $X$  on  $[a, b]$  relative to partition  $p_n$  are  $v_h(p_n) = \sum_{k=1}^n |h(t_k) - h(t_{k-1})|$  and  $v_X(p_n) = \sum_{k=1}^n \|X(t_k) - X(t_{k-1})\|$ , respectively, where  $\|X(t)\| = (E[|X(t)|^2])^{1/2}$  denotes the norm of  $X(t)$  in  $L^2$ . The corresponding total variations are  $v_h = \sup_p \{v_h(p)\}$  and  $v_X = \sup_p \{v_X(p)\}$ , where the sup is over all partitions  $p$  of  $[a, b]$ . If  $v_h < \infty$  and  $v_X < \infty$ ,  $h$  and  $X$  are said to be of bounded variation on  $[a, b]$ .

**Definition 3.21** Let  $p = (a = s_0 < s_1 < \cdots < s_m = b)$  and  $q = (a = t_0 < t_1 < \cdots < t_n = b)$  be partitions of  $[a, b]$ , let  $X$  be a real-valued process with correlation function  $r$ , and let  $\tilde{v}_r(p, q) = \sum_{k=1}^m \sum_{l=1}^n |r(s_k, t_l) - r(s_k, t_{l-1}) - r(s_{k-1}, t_l) +$

$r(s_{k-1}, t_{l-1})|$  denote the variation of  $r$  on  $[a, b] \times [a, b]$  relative to  $p$  and  $q$ . If  $\sup_{p,q} \{\tilde{v}_r(p, q)\} < \infty$ , then  $X$  is of bounded variation in the weak sense on  $[a, b]$ .

The following theorems, stated without proof, provide useful criteria for the existence of the integrals  $\int_a^b h(t) dX(t)$  and  $\int_a^b X(t) dh(t)$ .

**Theorem 3.14** *If  $X$  and  $Y$  are real-valued m.s. continuous processes and  $h : [a, b] \rightarrow \mathbb{R}$  is of bounded variation on  $[a, b]$ , then  $\int_a^b X(t) dh(t)$  exists,  $\|\int_a^b X(t) dh(t)\| \leq \max_{t \in [a, b]} \|X(t)\| v_h$ , and  $E[\int_a^b X(t) dt \int_a^b Z(t) dt] = \int_{[a, b]^2} E[X(u)Z(v)] du dv$ . ([5], Sect. 3.9.3.2).*

**Theorem 3.15** *If  $h : [a, b] \rightarrow \mathbb{R}$  is continuous and  $X$  is a real-valued m.s. differentiable process with m.s. continuous derivative  $\dot{X}$ , then  $\int_a^b h(t) dX(t) = \int_a^b h(t) \dot{X}(t) dt$  and both integrals exist ([5], Sect. 3.9.3.2).*

**Theorem 3.16** *Let  $h$  and  $k$  be real-valued functions defined on a bounded interval  $[a, b]$  of the real line,  $X, Y$  be real-valued processes, and  $\alpha, \beta \in \mathbb{R}$  be some constants. If the following integral exist, then ([5], Sect. 3.9.3.3).*

$$\begin{aligned}
 \int_a^b h(t) dX(t) &= h(t)X(t) \Big|_a^b - \int_a^b X(t) dh(t) \quad (\text{integration by parts}) \\
 \int_a^b [\alpha h(t) + \beta k(t)] dX(t) &= \alpha \int_a^b h(t) dX(t) + \beta \int_a^b k(t) dX(t) \quad (\text{linearity}) \\
 \int_a^b h(t) d[\alpha X(t) + \beta Y(t)] &= \alpha \int_a^b h(t) dX(t) + \beta \int_a^b h(t) dY(t) \quad (\text{linearity}) \\
 E \left[ \int_a^b h(t) dX(t) \right] &= \int_a^b h(t) dE[X(t)] \quad (\text{expectation and integration commute}) \\
 E \left[ \int_a^b X(t) dh(t) \right] &= \int_a^b E[X(t)] dh(t) \quad (\text{expectation and integration commute}).
 \end{aligned} \tag{3.24}$$

*Proof* For proof, see [2] (Sect. 2.3), [11] (Sect. 5.3), [8], and [5] (Sect. 3.9.3). Note that the latter formula in (3.24) follows by Fubini's theorem if  $X$  is measurable, that is, the function  $X : [a, b] \times \Omega \rightarrow \mathbb{R}$  is measurable from  $([a, b] \times \Omega, \mathcal{B}[a, b] \times \mathcal{F})$  to  $(\mathbb{R}, \mathcal{B})$ .  $\blacktriangle$

**Example 3.29** Let  $X$  be a real-valued m.s. continuous process and  $h : [0, \infty) \rightarrow \mathbb{R}$  be a function of bounded variation on compacts. Then  $Y(t) = \int_0^t X(s) dh(s)$ ,  $t \geq 0$ , is a m.s. differentiable process with mean  $E[Y(t)] = \int_0^t E[X(s)] dh(s)$  and correlation function  $E[Y(s)Y(t)] = \int_0^s \int_0^t E[X(u)X(v)] dh(u) dh(v)$ . The existence of  $Y$  and its second moment properties result from Theorem 3.14.  $\diamond$

### 3.6.4 Spectral Representation

Example 3.14 shows that superpositions of a finite number of harmonics with random amplitudes are real-valued, weakly stationary processes under some conditions. The spectral representation theorems in this section show that all weakly stationary random functions can be represented by sums of harmonics with random amplitudes.

**Theorem 3.17** *Let  $X$  be a complex-valued, weakly stationary, m.s. continuous process with spectral distribution  $S$ . There exists a complex-valued process  $Z$  with orthogonal increments, referred to as spectral process, defined up to an additive constant such that the m.s. integral*

$$X(t) = \int_{-\infty}^{\infty} e^{ivt} dZ(v) \quad (3.25)$$

*exists at any time  $t$ ,  $E[Z(v)] = 0$ ,  $E[|Z(v)|^2] = S(v)$  provided  $S(-\infty) = 0$ , and  $E[|dZ(v)|^2] = dS(v)$ . If  $S$  is absolutely continuous with respect to the Lebesgue measure, then  $dS(v) = s(v) dv$  so that  $E[|dZ(v)|^2] = s(v) dv$  ([8], Sect. 7.5).*

*Proof* The integral in (3.25) exists since the integrand  $e^{ivt}$  is continuous, the integrator  $Z(v)$  has the property

$$\begin{aligned} \tilde{v}_r(p, p) &= \sum_{k,l=1}^n |E[(Z(v_k) - Z(v_{k-1}))(Z(v_l) - Z(v_{l-1}))^*]| \\ &= \sum_{k=1}^n E[|Z(v_k) - Z(v_{k-1})|^2] = \sum_{k=1}^n (S(v_k) - S(v_{k-1})) = S(a) - S(-a), \end{aligned}$$

for an arbitrary partition  $p = (-a = v_0 < v_1 < \dots < v_n = a)$  of  $[-a, a]$ ,  $a > 0$ , and  $\lim_{a \rightarrow \infty} [S(a) - S(-a)] = S(\infty) < \infty$ , so that  $Z(v)$  is of bounded variation in the weak sense ([11], Theorem 2.29).  $\blacktriangle$

**Theorem 3.18** *Suppose  $X$  is as in Theorem 3.17 except it is real-valued. The integral representation,*

$$X(t) = \int_0^{\infty} [\cos(vt) dU(v) + \sin(vt) dV(v)], \quad (3.26)$$

*exists in m.s. at any time  $t$ , where the real-valued processes  $U$  and  $V$  have orthogonal increments with  $E[U(v)] = E[V(v)] = 0$ ,  $E[dU(v) dV(v')] = 0$ , and  $E[dU(v)^2] = E[dV(v)^2] = 2dS(v) = 2s(v) dv$  for all  $v, v' \geq 0$ , where the latter equality holds if  $S$  is absolutely continuous ([8], Sect. 7.6).*

**Example 3.30** Let  $X$  be a complex-valued process with spectral representation in (3.25) and let  $Y(t) = \sum_{k=1}^n c_k X(t + t_k)$ , where  $t_k \in \mathbb{R}$ ,  $c_k \in \mathbb{C}$ , and  $n \geq 1$  denote times, complex constants, and an integer. The spectral representation of  $Y$  is  $Y(t) = \int_{-\infty}^{\infty} e^{ivt} d\tilde{Z}(v)$ , where  $d\tilde{Z}(v) = h(v) dZ(v)$  and  $h(v) = \sum_{k=1}^n c_k e^{ivt_k}$  is the gain of the linear operator defining  $Y(t)$ .  $\diamond$

*Proof* We have  $Y(t) = \int_{-\infty}^{\infty} [\sum_{k=1}^n c_k e^{ivt_k}] e^{ivt} dZ(v) = \int_{-\infty}^{\infty} h(v) e^{ivt} dZ(v)$ , so that  $Y$  consists of a superposition of harmonics  $e^{ivt}$  with amplitudes  $h(v) dZ(v)$ . If  $\int_{-\infty}^{\infty} |h(v)|^2 dS(v) < \infty$ , then  $Y$  is in  $L^2$ .  $\blacktriangle$

**Example 3.31** If  $X$  is a real-valued weakly stationary process and  $\dot{X}$  exists in an m.s. sense, then

$$\dot{X}(t) = \int_{-\infty}^{\infty} i v e^{ivt} dZ(v) \quad (3.27)$$

with the notation in (3.25).  $\diamond$

*Proof* Let  $X^{(a)}(t) = \int_{-a}^a e^{ivt} dZ(v)$  and  $\dot{X}^{(a)}(t) = \int_{-a}^a i v e^{ivt} dZ(v)$  for  $a, \tau > 0$ . Straightforward calculations show that  $E[(X^{(a)}(t + \tau) - X^{(a)}(t))/\tau - \dot{X}^{(a)}(t)]^2$  converges to 0 as  $\tau \rightarrow 0$  for any  $a > 0$ . The formula in (3.27) follows by letting  $a \rightarrow \infty$  in the latter expectation.  $\blacktriangle$

**Theorem 3.19** Let  $X$  be an  $\mathbb{R}^d$ -valued, weakly stationary, and m.s. continuous process with spectral density functions  $\{s_{k,l}\}$ ,  $k, l = 1, \dots, d$ . The integral representation,

$$X(t) = \int_0^\infty [\cos(vt) dU(v) + \sin(vt) dV(v)] \quad (3.28)$$

exists in m.s. at any time, where  $U$  and  $V$  are  $\mathbb{R}^d$ -valued processes with mean zero and orthogonal increments such that

$$\begin{aligned} E[dU_k(v) dU_l(v')] &= E[dV_k(v) dV_l(v')] = \delta(v - v') g_{k,l}(v) dv \quad \text{and} \\ E[dU_k(v) dV_l(v')] &= -E[dV_k(v) dU_l(v')] = \delta(v - v') h_{k,l}(v) dv, \end{aligned} \quad (3.29)$$

where  $g_{k,l}(v) = s_{k,l}(v) + s_{k,l}(-v)$  and  $h_{k,l}(v) = -i[s_{k,l}(v) - s_{k,l}(-v)]$ .

*Proof* We have

$$\begin{aligned} r_{k,l}(\tau) &= E[X_k(t + \tau) X_l(t)] = \int_{\mathbb{R}} e^{iv\tau} s_{k,l}(v) dv \\ &= \int_0^\infty \left[ (s_{k,l}(v) + s_{k,l}(-v)) \cos(v\tau) + i(s_{k,l}(v) - s_{k,l}(-v)) \sin(v\tau) \right] dv \end{aligned}$$

by (3.16), so that  $r_{k,l}(\tau) = \int_0^\infty [g_{k,l}(v) \cos(v\tau) - h_{k,l}(v) \sin(v\tau)] dv$ . The expressions of  $r_{k,l}$  yield the second moment properties of  $U$  and  $V$  in (3.29).  $\blacktriangle$

**Example 3.32** Let  $Z$  be a complex-valued random field defined on  $\mathbb{R}^{d'}$  that has orthogonal increments with  $E[|Z(I)|] = 0$  and  $E[|Z(I)|^2] = S(I)$ , where  $S$  is a measure on  $\mathbb{R}^{d'}$  such that  $S(-\infty, \dots, -\infty) = 0$  and  $I$  is a Borel set in  $\mathbb{R}^{d'}$ . Then  $X(t) = \int_{\mathbb{R}^{d'}} \exp(it'v) dZ(v)$  is a weakly stationary random field with  $E[|X(t)|] = 0$  and  $E[|X(t)|^2] = \int_{\mathbb{R}^{d'}} dS(v)$ .  $\diamond$

*Proof* Consider the sequence  $Y_n(a) = \sum_{q=1}^n g(v_q)Z(I_q(a))$ , where  $g : \mathbb{R}^{d'} \rightarrow \mathbb{C}$  satisfies the condition  $\int_{\mathbb{R}^{d'}} |g(v)|^2 dS(v) < \infty$ ,  $\{I_q(a), q = 1, \dots, n\}$  denotes a partition of  $(-a, a]^{d'}$ ,  $a > 0$ , such that  $\lambda(I_q(a)) \rightarrow 0$  as  $n \rightarrow \infty$ , and  $v_q \in I_q(a)$ . The m.s. integral  $\int_{\mathbb{R}^{d'}} g(v) dZ(v)$  is defined as the limit of the Cauchy sequence  $\{Y_n(a)\}$  in  $L^2$  as  $n \rightarrow \infty$  and  $a \rightarrow \infty$ . The mean and variance of  $\int_{\mathbb{R}^{d'}} g(v) dZ(v)$  are 0 and  $\int_{\mathbb{R}^{d'}} |g(v)|^2 dS(v)$ . The field  $X(t)$  is a special case of  $\int_{\mathbb{R}^{d'}} g(v) dZ(v)$ .  $\blacktriangle$

**Theorem 3.20** *Let  $X$  be a real-valued, m.s. continuous, weakly homogeneous random field defined on  $\mathbb{R}^{d'}$ . The integral representation,*

$$X(t) = \int_{\mathbb{R}^{d'}} e^{it'v} dZ(v), \quad (3.30)$$

*exists in m.s. at any  $t \in \mathbb{R}^{d'}$ , where  $Z$  is a complex valued random field with orthogonal increments defined on  $\mathbb{R}^{d'}$  such that  $E[dZ(v)] = 0$ ,  $E[|dZ(v)|^2] = dS(v)$ , and  $S$  is defined by (3.17). The correlation function of  $X$  is  $r(s, t) = \int_{\mathbb{R}^{d'}} e^{i(s-t)v} dS(v)$ , ([2], Theorem 2.4.1).*

### 3.6.5 Karhunen–Loève Expansion

Let  $X(t)$ ,  $t \in I$ , be a complex-valued random function with mean 0, finite variance, and correlation function  $r(s, t) = E[X(s)X(t)^*]$ , where  $I \subset \mathbb{R}^{d'}$  is a compact set. The integral equation,

$$\int_I r(s, t)\phi(t) dt = \lambda\phi(s), \quad s \in I, \quad (3.31)$$

has the form  $T\phi = \lambda\phi$ , where  $T : L^2(I) \rightarrow L^2(I)$  is a linear operator defined by  $T\phi(s) = \int_I r(s, t)\phi(t) dt$  and  $L^2(I)$  is the Hilbert space of square integrable functions on  $I$  with the inner product  $\langle \phi, \xi \rangle = \int_I \phi(t)\xi(t)^* dt$ .

**Definition 3.22** A non-zero number  $\lambda$  for which there exists a function  $\phi$  satisfying both (3.31) and the integrability condition  $\int_I |\phi(t)|^2 dt < \infty$  is called an eigenvalue. The function  $\phi$  is the eigenfunction of  $\lambda$ .

**Theorem 3.21** *The eigenvalues  $\lambda$  of  $T\phi = \lambda\phi$  are real-valued and the eigenfunctions corresponding to distinct eigenvalues are orthogonal, that is,  $\int_I \phi(t)\varphi(t)^* dt = 0$  for any pair  $(\phi, \varphi)$  of eigenfunctions associated with distinct eigenvalues.*

*Proof* This follows from Theorem B.58 since  $T$  is a self-adjoint operator. The eigenfunctions  $\phi$  can be scaled to have unit norm, that is,  $\int_I |\phi(t)|^2 dt = 1$ .  $\blacktriangle$

Note that (3.31) has at least an eigenvalue and an eigenfunction, the collection of eigenvalues of (3.31) is at most countable, and for each eigenvalue  $\lambda_k$  there exists at most a finite number of linearly independent eigenfunctions (Sect. B.4.5, [2], Sect. 3.3 [12], Appendix 2, and [13], Sect. 6.2).

**Theorem 3.22** *If  $r$  is square integrable and continuous in  $I \times I$ , then  $X$  admits the representation*

$$X(t) = \text{l.i.m.}_{n \rightarrow \infty} \sum_{k=1}^n \lambda_k^{1/2} X_k \phi_k(t), \quad t \in I, \quad \text{where}$$

$$X_k = \lambda_k^{-1/2} \int_I X(t) \phi_k(t)^* dt, \quad E[X_k] = 0, \quad \text{and} \quad E[X_k X_l^*] = \delta_{kl}. \quad (3.32)$$

Moreover, the Mercer theorem holds and gives  $r(t, s) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(s)^*$ .

*Proof* The mean and correlation functions of  $X^{(n)}(t) = \sum_{k=1}^n \lambda_k^{1/2} X_k \phi_k(t)$  are  $E[X^{(n)}(t)] = 0$  and  $E[X^{(n)}(s)(X^{(n)}(t))^*] = \sum_{k=1}^n \lambda_k \phi_k(s) \phi_k(t)^*$  ([2], Sect. 3.3, [12], Appendix 2, [13], Sect. 6.2).  $\blacktriangle$

*Example 3.33* Let  $X$  be a real-valued, weakly stationary, m.s. periodic process with period  $T > 0$ , mean 0, and correlation function  $r(\tau) = E[X(t + \tau)X(t)]$ . The Karhunen–Loève expansion and the spectral representation of  $X$  coincide.  $\diamond$

*Proof* A process  $X$  is said to be m.s. periodic with period  $T$ , if its correlation function is periodic with period  $T$ , that is,  $r(\tau) = \sum_{k=-\infty}^{\infty} c_k \exp(ikv_0\tau)$ , where  $v_0 = 2\pi/T$  and  $c_k = c_{-k}$ . The spectral density of  $X$  is  $s(v) = \sum_{k=-\infty}^{\infty} c_k \delta(v - kv_0)$ , and has a countable number of frequencies. The solution of (3.31) for  $d' = 1$  and  $I = [0, T]$  are  $\lambda_k = T c_k$  and  $\phi_k(t) = \exp(ikv_0 t) / \sqrt{T}$ ,  $k = 0, \pm 1, \pm 2, \dots$ , so that  $X(t) = \text{l.i.m.}_{n \rightarrow \infty} X^{(n)}(t)$ , where  $X^{(n)}(t) = \sum_{k=-n}^n V_k \exp(ikv_0 t)$  and the random variables in the representation of  $X^{(n)}(t)$  are given by  $V_k = (1/T) \int_0^T X(t) \exp(-ikv_0 t) dt$ . The spectral representation-based expansion of  $X$  has the same functional form (Example 3.15). Related results can be found in [14].  $\blacktriangle$

*Example 3.34* Let  $X(t)$ ,  $t \in [-\xi, \xi]$ ,  $\xi > 0$ , be a real-valued stochastic process with mean zero and correlation function  $r(t, s) = (1/4)e^{-2\alpha|\tau|}$ , where  $\tau = t - s$  and  $\alpha > 0$ . The Karhunen–Loève representation for  $X$  in  $[-\xi, \xi]$  is

$$X(t) = \sum_{k=1}^{\infty} \left[ \sqrt{\lambda_k} X_k \phi_k(t) + \sqrt{\hat{\lambda}_k} \hat{X}_k \hat{\phi}_k(t) \right], \quad (3.33)$$

where

$$\lambda_k = \frac{1}{4\alpha(1 + b_k^2)}, \quad \phi_k(t) = \frac{\cos(2\alpha b_k t)}{\sqrt{\xi + \sin(4\alpha b_k \xi)/(4\alpha b_k)}},$$

$$\hat{\lambda}_k = \frac{1}{4\alpha(1 + \hat{b}_k^2)}, \quad \hat{\phi}_k(t) = \frac{\sin(2\alpha \hat{b}_k t)}{\sqrt{\xi - \sin(4\alpha \hat{b}_k \xi)/(4\alpha \hat{b}_k)}},$$

$b_k, \hat{b}_k$  are the solutions of  $b_k \tan(2\alpha \xi b_k) = 1$ ,  $\hat{b}_k \cot(2\alpha \xi \hat{b}_k) = 1$ , and the series in (3.33) converges in mean square. The eigenvalues and eigenfunctions in the representation of  $X$  are the solutions of the integral equation  $\int_{-\xi}^{\xi} e^{-2\alpha|t-s|} \phi(s) ds = 4\lambda \phi(t)$  ([12], Example 6-4.1, p. 99).  $\diamond$



*Example 3.35* The Karhunen–Loève representation of a real-valued process  $X$  defined on  $[0, 1]$  with mean zero and correlation function  $r(t, s) = t \wedge s$  is

$$X(t) = \text{l.i.m.}_{n \rightarrow \infty} \frac{\sqrt{2}}{\pi} \sum_{k=0}^n \frac{\sin(k + 1/2)\pi t}{k + 1/2} X_k, \quad t \in [0, 1], \quad (3.34)$$

where  $X_0, X_1, \dots$  are uncorrelated random variables with means  $E[X_k] = 0$  and variances  $E[X_k^2] = 1$ .  $\diamond$

*Proof* The integral equation (3.31) becomes  $\int_0^t s\phi(s) ds + t \int_t^1 \phi(s) ds = \lambda\phi(t)$ ,  $t \in (0, 1)$ . The first and second derivatives of this equation with respect to  $t$  are  $\int_t^1 \phi(s) ds = \lambda\phi'(t)$  and  $\lambda\phi''(t) + \phi(t) = 0$ , respectively. The boundary conditions for the latter equation are  $\phi(0) = 0$  and  $\phi'(1) = 0$ , and result from (3.31) and its first derivative at  $t = 0$  and  $t = 1$ , respectively. The solution of  $\lambda\phi''(t) + \phi(t) = 0$  with these boundary conditions gives the eigenvalues and eigenfunctions for the Karhunen–Loève representation of  $X$ .

The representation in (3.34) applies to both Brownian motion and compound Poisson processes since their correlation functions coincide with that of  $X$ . This shows that the Karhunen–Loève representation cannot distinguish between processes with the same second moment properties.  $\blacktriangle$

*Example 3.36* Let  $W(t)$ ,  $t \in [a, b]$ , be a white noise defined heuristically as a process with mean zero and correlation function  $E[W(t)W(s)] = \gamma\delta(t - s)$ , where  $s, t \in [a, b]$ ,  $\gamma > 0$ . This definition of  $W$  is frequently used in engineering applications [10]. The Karhunen–Loève representation of  $W$  is  $W(t) = \sqrt{\gamma} \sum_{k=1}^{\infty} W_k \phi_k(t)$ , where  $W_k$  are random variables with  $E[W_k] = 0$  and  $E[W_k W_l] = \delta_{kl}$  and  $\{\phi_k\}$  is an arbitrary family of orthonormal functions spanning  $L^2[a, b]$ . The representation follows from (3.31), that takes the form  $\gamma\phi(t) = \lambda\phi(t)$  for white noise processes. The calculations in this example are formal since  $W$  does not exist in the second moment sense.  $\diamond$

### 3.7 Classes of Stochastic Processes

We have seen that the concept of stationarity need not be specialized for stochastic processes and random fields. Similarly, it is no need to provide distinct definitions for Gaussian, translation, and ergodic stochastic processes and random fields. However, Markov and independent increment random functions in time and space differ significantly. This section does not discuss Markov random fields. Useful information on this topic can be found in [2] (Chap. 8).

### 3.7.1 Gaussian Random Functions

**Definition 3.23** An  $\mathbb{R}^d$ -valued random function  $X(t)$ ,  $t \in \mathbb{R}^{d'}$ , is said to be Gaussian if its finite dimensional distributions are Gaussian.

Note that weakly stationary Gaussian functions are stationary and linear transformations of Gaussian functions are Gaussian (Exercise 3.11). Also, the processes  $Z$  and  $(U, V)$  in the spectral representations given by (3.25) and (3.26) and the random variables  $X_k$  of the Karhunen–Loève representation in (3.32) are Gaussian.

*Example 3.37* Let  $X(t)$ ,  $t \in I$ , be a Gaussian function with mean 0, where  $I$  is a closed set in  $\mathbb{R}^{d'}$ . The random variables  $\{X_k\}$  in the Karhunen–Loève expansion of  $X(t)$  given by (3.32) are independent Gaussian variables with means  $E[X_k] = 0$  and covariances  $E[X_k X_l] = \delta_{kl}$ .  $\diamond$

*Proof* The random variables  $X_k$  in (3.32) are stochastic integrals defined as m.s. limits of the sequences  $S_{X,h}(p_n)$  in (3.23). These sequences are Gaussian variables as linear transformations of Gaussian variables, that is, values of  $X(t)$  at discrete times, with characteristic functions  $\varphi_{S_{X,h}(p_n)}(u) = \exp(-u^2 \text{Var}[S_{X,h}(p_n)]/2)$ . Since  $\exp(i\xi)$  is Lipschitz, that is, there is a constant  $\alpha > 0$  such that  $|\exp(i\xi') - \exp(i\xi'')| \leq \alpha|\xi' - \xi''|$ , we have

$$|e^{iuS_{X,h}(p_n)} - e^{iu \int_I X(t)\phi_k(t)^* dt}| \leq \alpha|u| |S_{X,h}(p_n) - \int_I X(t)\phi_k(t)^* dt|,$$

so that  $\exp(iuS_{X,h}(p_n)) \rightarrow \exp(iu \int_I X(t)\phi_k(t)^* dt)$  in m.s., and

$$\begin{aligned} E\left[\exp\left(iu \int_I X(t)\phi_k(t)^* dt\right)\right] &= \lim_{n \rightarrow \infty} E[\exp(iuS_{X,h}(p_n))] \\ &= \lim_{n \rightarrow \infty} \exp(-u^2 \text{Var}[S_{X,h}(p_n)]/2) = \exp\left(-u^2 \text{Var}\left[\int_I X(t)\phi_k(t)^* dt\right]/2\right) \end{aligned}$$

by Theorem 3.11, so that  $X_k$  is a Gaussian variable.  $\blacktriangle$

### 3.7.2 Translation Random Functions

**Definition 3.24** Let  $G(t)$ ,  $t \in \mathbb{R}^{d'}$ , be an  $\mathbb{R}^d$ -valued stationary Gaussian function and let  $h_i : \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $i = 1, \dots, d$ , be measurable functions. The  $\mathbb{R}^d$ -valued random function  $X(t)$  with coordinates  $X_i(t) = h_i(G(t))$ ,  $t \in \mathbb{R}^{d'}$ , is referred to as translation random function.

It is common in applications to define the coordinates of  $X(t)$  by  $X_i(t) = h_i(G_i(t))$ ,  $i = 1, \dots, d$ , where  $h_i : \mathbb{R} \rightarrow \mathbb{R}$  are continuous functions, so that  $X_i(t)$  are memoryless transformations of the coordinates  $\{G_i(t)\}$  of  $G(t)$ .

**Theorem 3.23** *If  $X(t), t \in \mathbb{R}^{d'}$ , is an  $\mathbb{R}^d$ -valued translation random function defined by  $X_i(t) = h_i(G_i(t))$ ,  $i = 1, \dots, d$ , where  $h_i = F_i^{-1} \circ \Phi : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\{F_i\}$  denote distributions with finite variances,  $\Phi$  is the distribution of  $N(0, 1)$ , and the coordinates of  $G(t)$ ,  $t \in \mathbb{R}^{d'}$ , are real-valued Gaussian functions with mean 0 and variance 1, then*

- (1) *If  $G$  is stationary so is  $X$ ,*
- (2)  $P(X_i(t) \leq x_i) = F_i(x_i)$ ,  $i = 1, \dots, d$ ,
- (3)  $E[X_i(t)] = \int_{\mathbb{R}} h_i(u) \phi(u) du$ ,
- (4)  $E[X_i(s)X_j(t)] = \int_{\mathbb{R}^2} h_i(u)h_j(v) \phi(s, t; \rho_{ij}(u, v)) du dv$   
 $= E[h_i(G_i(s))h_j(G_j(t))]$ ,
- (5)  $|\zeta_{ij}(s, t)| \leq |\rho_{ij}(s, t)|$ ,  $s, t \in \mathbb{R}^{d'}$ , and
- (6) *Given  $(F_i(\cdot), \zeta_{ij}(\cdot, \cdot))$ ,  $\exists G_i$  such that  $X_i(t) = h_i(G_i(t))$  if and only if  $\rho_{ij}(\cdot, \cdot)$  satisfying (4) are correlation functions,  $i, j = 1, \dots, d$ ,*

(3.35)

where  $\phi(\cdot)$  is the density of  $N(0, 1)$ ,  $\phi(\cdot, \cdot; \rho)$  is the joint density of a standard bivariate Gaussian vector with correlation coefficient  $\rho$ ,  $\zeta_{ij}(s, t) = (E[X_i(s)X_j(t)] - E[X_i(s)]E[X_j(t)]) / (\text{Var}[X_i(s)]\text{Var}[X_j(t)])^{1/2}$  and  $\text{Var}[X_i(s)] = E[X_i(s)^2] - E[X_i(s)]^2$ .

*Proof* Since  $P(X_i(t_1) \leq x_1, \dots, X_i(t_n) \leq x_n) = P(G_i(t_1) \leq y_1, \dots, G_i(t_n) \leq y_n)$ ,  $y_i = \Phi^{-1} \circ F_i(x_i)$ , then  $X$  is stationary if  $G$  has this property. The marginal distribution of  $X_i$  is  $P(X_i(u) \leq \xi) = P(G_i(u) \leq \Phi^{-1} \circ F_i(\xi)) = F_i(\xi)$ . Formulas in (3) and (4) result from the definitions of  $X$  and of the joint distribution of  $(G_i(s), G_j(t))$ . The inequality in (5) follows from a maximal property of bivariate Gaussian distributions [15]. For (6), first note that  $\rho_{ii}(\tau) = 0; 1$  implies  $\zeta_{ii}(\tau) = 0; 1$ . The correlation coefficient  $\zeta_i^* = \zeta_{ii}(\tau)$  corresponding to  $\rho_{ii}(\tau) = -1$ , can be calculated from

$$\zeta_i^* = \frac{E[h_i(G_i)h_i(-G_i)] - E[h_i(G_i)]^2}{E[h_i(G_i)^2] - E[h_i(G_i)]^2}, \quad (3.36)$$

and has the property  $\zeta_i^* \geq -1$ . The value  $\zeta_i^* = -1$  can be reached if, for example,  $h_i$  is an odd function, in which case  $E[h_i(G_i)] = 0$  and  $E[h_i(G_i)h_i(-G_i)] = -E[h_i(G_i)h_i(G_i)]$ . A theorem by Price ([4], Sect. 3.1.1) gives the relationship

$$\frac{\partial^{(k)} \zeta_{ii}(\tau)}{\partial \rho_{ii}(\tau)^{(k)}} = \frac{1}{\text{Var}[X_i(t)]} E[h_i^{(k)}(G_i(t + \tau))h_i^{(k)}(G_i(t))], \quad i = 1, \dots, d, \quad (3.37)$$

so that, since  $h_i'(\cdot) > 0$ ,  $\zeta_{ii}(\tau)$  is an increasing function of  $\rho_{ii}(\tau) \in [-1, 1]$  taking values in  $[\zeta_i^*, 1]$ .

Translation functions do not exist for pairs  $(\{F_i\}, \{\zeta_{ij}\})$  with  $\zeta_{ii}$  taking values outside the range  $[\zeta_i^*, 1]$ . The requirement  $\zeta_{ii}(\tau) \in [\zeta_i^*, 1]$ ,  $\tau \in \mathbb{R}$ , is necessary but not sufficient for the existence of translation functions. The existence of translation functions also requires that the images  $\rho_{ii}$  of  $\zeta_{ii}$  be positive definite. Additional considerations on the existence of translation random functions can be found in [4] (Sect. 3.1) and [16]. ▲

*Example 3.38* Let  $X(t) = G(t)^3$ ,  $t \in \mathbb{R}$ , where  $G$  is a stationary Gaussian process with mean 0, variance 1, and correlation function  $\rho(\tau) = E[G(t + \tau)G(t)]$ . The scaled covariance function of  $X$  is  $\zeta(\tau) = \rho(\tau)(3 + 2\rho(\tau))/5$ , a result that can be obtained by direct calculations. We have  $\zeta^* = -1$  in agreement with a previous observation regarding odd transformations. ◇

*Example 3.39* Set  $X(t) = \exp(G(t))$ ,  $t \in \mathbb{R}$ , where  $G(t)$  is as in Example 3.38. Then  $\zeta(\tau) = (1 - \exp(\rho(\tau)))/(1 - \exp(1))$ , so that  $\zeta^* \simeq -0.3679$ . As previously indicated,  $\zeta(\tau) = 0$  and 1 for  $\rho(\tau) = 0$  and 1, respectively. ▲

*Example 3.40* Let  $X(t), t \geq 0$ , be a real-valued translation process defined by  $X(t) = F^{-1} \circ \Phi(G(t))$ , where  $G(t)$  denotes a real-valued stationary Gaussian process with mean 0, variance 1, and correlation function  $\rho(\tau) = E[G(t + \tau)G(t)]$ , where  $F$  is an absolutely continuous distribution function with density  $f$ . The finite dimensional densities of  $X$  are

$$\begin{aligned} & f(x_1, \dots, x_n; t_1, \dots, t_n) \\ &= [(2\pi)^n \det(\rho)]^{-1/2} \prod_{p=1}^n \frac{f(x_p)}{\phi(y_p)} \exp\left(-\frac{1}{2}y' \rho^{-1}y\right), \end{aligned} \quad (3.38)$$

where  $\rho = \{E[G(t_p)G(t_q)]\}$ ,  $y_p = \Phi^{-1} \circ F(x_p)$ ,  $p, q = 1, \dots, n$ ,  $\phi$  is the density of  $N(0, 1)$ , and  $y = (y_1, \dots, y_n)$ . ◇

*Proof* The finite dimensional distribution  $F$  of order  $n$  of  $X$  is the probability of the event  $\{G(t_1) \leq y_1, \dots, G(t_n) \leq y_n\}$ , and  $f_n(x_1, \dots, x_n; t_1, \dots, t_n)$  results by differentiation. Alternatively,  $f$  can be obtained from the density of the Gaussian vector  $(G(t_1), \dots, G(t_n))$  and the change of variable defining  $X$ . ▲

### 3.7.3 Ergodic Random Functions

Useful information on ergodic random functions can be found in [2] (Sect. 6.5), [7] (Sect. 2.6), and [17] (Chap. 13). This section only provides a brief introduction on ergodicity.

**Definition 3.25** An  $\mathbb{R}^d$ -valued stationary random function  $X(t)$ ,  $t \in \mathbb{R}^{d'}$ , is called ergodic if temporal/spatial averages of a particular sample coincide with  $E[X(t)]$ , that is, temporal/spatial and ensemble averages coincide.

We give without proof two results for stationary Gaussian random functions providing simple criteria for assessing whether a Gaussian function is or is not ergodic.

Note that translation random functions defined by continuous mappings of ergodic Gaussian functions are ergodic.

**Theorem 3.24** *A stationary Gaussian random function is ergodic if and only if its spectral distribution function is continuous everywhere ([2], Theorem 6.5.3).*

**Theorem 3.25** *A stationary Gaussian random function is ergodic if its correlation function  $r$  satisfies the condition  $r(\tau) \rightarrow 0$  as  $\|\tau\| \rightarrow \infty$  ([2], Theorem 6.5.4).*

A heuristic interpretation of the latter theorem is that for an ergodic Gaussian function  $X$  the random variables  $X(t)$  and  $X(t + \tau)$  are approximately independent for a sufficiently large  $\|\tau\|$ .

*Example 3.41* Let  $X_k = Y + Y_k$ ,  $k = 1, 2, \dots$ , be a time series, where  $\{Y_k\}$  are iid real-valued random variables independent of random variable  $Y$  and  $Y_1, Y \in L^2$ . The time series  $\{X_k\}$  is stationary but is not ergodic since its temporal mean  $(1/n) \sum_{k=1}^n X_k = Y + (1/n) \sum_{k=1}^n Y_k$  converges a.s. to  $Y + E[Y_1]$  as  $n \rightarrow \infty$  by the strong law of large numbers (Example 2.32), so that it depends on the particular sample of  $Y$ .  $\diamond$

It is common in applications to consider weaker definitions for ergodicity, such that ergodicity in the mean and mean square. For example, we say that a real-valued process  $X$  is ergodic in the mean if  $\lim_{\tau \rightarrow \infty} \int_{-\tau}^{\tau} X(t) dt / (2\tau) = E[X(t)]$  with probability one.

*Example 3.42* A real-valued weakly stationary process  $X$  with mean  $\mu$  and covariance function  $c(h) = E[(X(t+h) - \mu)(X(t) - \mu)]$  is ergodic in the mean if

$$\lim_{\tau \rightarrow \infty} \text{Var}[X_\tau] = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{-\tau}^{\tau} (1 - |\alpha|/\tau) c(\alpha) d\alpha = 0. \quad (3.39)$$

Note that  $X$  is ergodic in the mean if its covariance function is absolutely integrable since  $|\int_{-\tau}^{\tau} (1 - |\alpha|/\tau) c(\alpha) d\alpha| \leq \int_{-\tau}^{\tau} |c(\alpha)| d\alpha < \int_{-\infty}^{\infty} |c(\alpha)| d\alpha$ , and the latter integral is bounded by assumption. For example,  $X$  with  $c(h) = \exp(-\lambda|h|)$ ,  $\lambda > 0$ , is ergodic in the mean since  $\int_{-\infty}^{\infty} |c(\alpha)| d\alpha = 2/\lambda$ .  $\diamond$

*Proof* Let  $X_\tau = (1/\tau) \int_{-\tau/2}^{\tau/2} X(t) dt$  and  $\tilde{X}_\tau = (1/\tau) \int_{-\tau/2}^{\tau/2} \tilde{X}(t) dt$ , where  $\tilde{X}(t) = X(t) - \mu$ . Note that  $X_\tau$  is an unbiased estimator for the mean of  $X(t)$  since  $E[X_\tau] = \mu$ . The variance of  $X_\tau$  is

$$\begin{aligned} \text{Var}[X_\tau] &= E[\tilde{X}_\tau^2] = \frac{1}{\tau^2} \int_{-\tau/2}^{\tau/2} \int_{-\tau/2}^{\tau/2} c(t-s) ds dt \\ &= \frac{1}{\tau^2} \left[ \int_{-\tau}^0 c(\alpha) d\alpha \int_{-\tau/2-\alpha}^{\tau/2} d\beta + \int_0^{\tau} c(\alpha) d\alpha \int_{-\tau/2}^{\tau/2-\alpha} d\beta \right] \\ &= \frac{1}{\tau^2} \left[ \int_{-\tau}^0 c(\alpha)(\tau + \alpha) d\alpha + \int_0^{\tau} c(\alpha)(\tau - \alpha) d\alpha \right] \\ &= \frac{1}{\tau^2} \int_{-\tau}^{\tau} (\tau - |\alpha|) c(\alpha) d\alpha = \frac{1}{\tau} \int_{-\tau}^{\tau} (1 - |\alpha|/\tau) c(\alpha) d\alpha \end{aligned}$$

by the change of variable ( $\alpha = t - s, \beta = s$ ). If  $\text{Var}[X_\tau] \rightarrow 0$  as  $\tau \rightarrow \infty$ , a single sample of  $X$  suffices to find its mean provided the sample is infinitely long. ▲

### 3.7.4 Markov Random Functions

Markov chains and processes are defined and illustrated by examples. Markov random fields differ significantly from Markov chains and processes since time has a natural flow from past to future while space does not have a preferred evolution. We do not discuss Markov fields; useful information on this class of random functions can be found in [2] (Appendix), [18–21], and [7] (Sect. 8.4).

#### 3.7.4.1 Markov Chains

Let  $X$  be a real-valued Markov chain taking values in a finite set  $S = \{1, 2, \dots, q\}$  called state space. We refer to  $X$  as discrete time, discrete state Markov process or discrete state Markov chain. Denote by  $X_n, n = 0, 1, \dots$ , the state  $X$  at time  $n$ ,  $\pi(n) = (\pi_1(n) = P(X_n = 1), \dots, \pi_q(n) = P(X_n = q))'$  a  $q$ -dimensional vector, and  $p(m, n) = \{p_{ij}(m, n) = P(X_n = j \mid X_m = i), i, j = 1, \dots, q\}, n \geq m \geq 0$ , a  $(q, q)$ -matrix of conditional probabilities ([22], Chaps. 5 and 6, and [23], Chap. 2).

**Definition 3.26** If  $p(m, n)$  has the properties  $p_{ij}(m, n) = P(X_n = j \mid X_m = i) \in [0, 1]$  for all  $i, j = 1, \dots, q$  and  $\sum_{j=1}^q p_{ij}(n, n+1) = 1$  at all  $n \geq 0$  and states  $i \in S$ , it is called transition probability matrix. The second condition means that  $X_{n+1} \in S$  with probability 1 conditional on  $\{X_n = i\}, i \in S$ .

**Theorem 3.26** If  $\{X_n, n = 0, 1, \dots\}$  is a Markov chain with values in  $S = \{1, 2, \dots, q\}$ , then

$$\begin{aligned} \pi(n)' &= \pi(m)' p(m, n), \quad n \geq m \geq 0, \quad \text{and} \\ p(n, k) &= p(k, m) p(m, n), \quad n \geq m \geq k \geq 0. \end{aligned} \quad (3.40)$$

The second relationship is the Chapman–Kolmogorov equation.

*Proof* Conditional on  $\{X_m = i\}$ , we have  $\{X_n = j\}$  with probability  $p_{ij}(m, n)$  so that  $P(X_n = j) = \sum_{i=1}^q p_{ij}(m, n) P(X_m = i)$ , that is, the first relationship. The equalities  $P(X_n = j, X_m = l, X_k = i) = P(X_n = j \mid X_m = l, X_k = i) P(X_m = l, X_k = i) = P(X_n = j \mid X_m = l) P(X_m = l, X_k = i)$  give  $p_{ij}(k, n) = \sum_{l=1}^q p_{lj}(m, n) p_{il}(k, m)$  by summation over all  $l \in S$  or  $p(k, n) = p(k, m) p(m, n)$ . ▲

**Definition 3.27** A Markov chain with transition probability matrix  $p(m, n), n \geq m \geq 0$ , is said to be homogeneous if  $p(m, n) = p(0, n - m)$ , that is, its transition

matrix depends on the time lag  $n - m$  rather than  $m$  and  $n$ . By abuse of notation, we use  $p(n - m)$  for  $p(0, n - m)$  and  $p = p(1)$  for  $p(0, 1)$ .

**Theorem 3.27** Let  $\{X_n, n = 0, 1, \dots\}$  be a homogeneous Markov chain taking values in  $S = \{1, 2, \dots, q\}$  with transition probability matrix  $p = p(1)$ . Then (Exercise 3.13)

$$\begin{aligned} p(n) &= p^n, \quad n = 1, 2, \dots, \\ p_{ij} &\geq 0, i, j \in S, \quad \text{such that } \sum_{j=1}^q p_{ij} = 1, \quad j \in S, \quad \text{and} \\ p(n + m) &= p(n)p(m), \quad n, m \geq 0. \end{aligned} \tag{3.41}$$

**Example 3.43** Let  $X_0 = 0$  and  $X_n = \sum_{k=1}^n Y_k$ ,  $n \geq 1$ , where  $Y_k$  are iid random variables with probabilities  $q_r = P(Y_1 = r)$ , where  $r \geq 0$  is an integer. Since  $P(X_{n+1} = j \mid X_0, X_1, \dots, X_n) = P(Y_{n+1} = j - X_n) = q_{j-X(n)}$ ,  $\{X_n, n = 0, 1, \dots\}$  is a Markov chain with transition probabilities  $p_{ij} = P(X_{n+1} = j \mid X_n = i) = q_{j-i}$ .  $\diamond$

**Definition 3.28** Let  $\{X_n, n = 0, 1, \dots\}$  be a Markov chain with state space  $S$ . Denote by  $T_i$  the time of the first visit to state  $i$  and by  $N_i$  the total number of visits to  $i$ . The state  $i$  is recurrent if  $P(T_i < \infty) = 1$ ; otherwise, if  $P(T_i = +\infty) > 0$ , state  $i$  is said to be transient. A recurrent state  $i$  is null if  $E[T_i] = \infty$ ; otherwise, is called non-null. A recurrent state is periodic with period  $\tau$  if  $\tau \geq 2$  is the largest integer for which  $P(T_i = n\tau \text{ for some } n \geq 1) = 1$ ; otherwise, if there is no such  $\tau \geq 2$ ,  $i$  is called aperiodic.

**Definition 3.29** A collection of states  $C \subseteq S$  is closed if no state  $j \notin C$  can be reached from a state in  $C$ , that is,  $P(X_{n+1} \notin C \mid X_n = i) = 0$  for all  $i \in C$ . If  $C = \{i\}$  consists of a single state, then  $i$  is called an absorbing state. A closed  $C$  is irreducible if no proper subset of it is closed. A Markov chain is called irreducible if its only closed subset is the set of all states.

**Example 3.44** Let  $\{X_n, n = 0, 1, \dots\}$  be homogeneous Markov chain with transition probability matrix

$$p = \begin{bmatrix} 0.50 & 0.25 & 0.25 & 0.00 \\ 0.00 & 0.50 & 0.50 & 0.00 \\ 0.50 & 0.00 & 0.50 & 0.00 \\ 0.00 & 0.50 & 0.50 & 0.00 \end{bmatrix}.$$

State 4 is transient since it cannot be reached. The collections of states  $\{1, 2, 3, 4\}$  and  $\{1, 2, 3\}$  are closed since, for example,  $X_n \in \{1, 2, 3\}$  will never leave this set of states. The states 1, 2, and 3 are not periodic since they can be accessed at every time step.  $\diamond$

When dealing with homogeneous Markov chains we may be interested in their long-term behavior, that is, the probability law of  $X_n$  as  $n \rightarrow \infty$ . If the probability law  $\pi(n)$  of  $X_n$  becomes invariant as  $n \rightarrow \infty$ , then  $\pi_{\text{st}} = \lim_{n \rightarrow \infty} \pi(n)$  exists and is called stationary distribution. Conditions for the existence of the stationary distribution can be found in [23] (Sect.2.12). The following theorem provides a method for calculating stationary distributions.

**Theorem 3.28** *Let  $\{X_n, n = 0, 1, \dots\}$  be a homogeneous Markov chain with state space  $S = \{1, 2, \dots, q\}$  and transition probability matrix  $p = p(1)$  that is irreducible and aperiodic. Then the stationary probability  $\pi_{\text{st}}$  is the solution of  $\pi'_{\text{st}} = (1, \dots, 1)(I - p - I_q)^{-1}$ , where  $I$  is the identity matrix and  $I_q$  is a  $(q, q)$ -matrix with all entries equal to 1 ([23], Proposition 2.14.1).*

**Example 3.45** Let  $\{X_n, n = 0, 1, \dots\}$  be a homogeneous Markov chain with transition probability matrix

$$p = \begin{bmatrix} 0 & 1/3 & 2/3 \\ 1/3 & 0 & 2/3 \\ 1 & 0 & 0 \end{bmatrix}.$$

The stationary distribution of the chain has been calculated from  $\pi'_{\text{st}} = (1, \dots, 1)(I - p - I_q)^{-1}$  and is  $\pi_{\text{st}}(1) = 0.45$ ,  $\pi_{\text{st}}(2) = 0.15$ , and  $\pi_{\text{st}}(3) = 0.40$ .  $\diamond$

### 3.7.4.2 Markov Processes

**Definition 3.30** Let  $X(t)$ ,  $t \geq 0$ , be a real-valued stochastic process. If for every integer  $n \geq 1$  and times  $0 \leq t_1 < \dots < t_n$  the conditional random variables  $X(t_n) \mid X(t_{n-1})$  and  $(X(t_{n-2}), \dots, X(t_1)) \mid X(t_{n-1})$  are independent, then  $X$  is a Markov process, that is, past  $(X(t_{n-2}), \dots, X(t_1))$  is independent of future  $X(t_n)$  conditional on present  $X(t_{n-1})$ . The definition extends directly to vector-valued processes.

The condition  $E[g(X(t+\tau)) \mid \mathcal{F}_t] = E[g(X(t+\tau)) \mid X(t)]$ , required to be satisfied for arbitrary  $t, \tau \geq 0$  and Borel functions  $g : \mathbb{R} \rightarrow \mathbb{R}$ , provides an alternative definition for Markov processes ([24], p. 156), where  $\mathcal{F}_t = \sigma(X(s), 0 \leq s \leq t)$ .

**Definition 3.31** The transition density of  $X$  is the density of the conditional random variable  $X(t) \mid X(s)$ ,  $0 \leq s < t$ .

**Theorem 3.29** *The finite dimensional densities of a Markov process  $X$  can be obtained from its transition density and its marginal density at the initial time. Moreover, the random variables  $X(t_n) \mid (X(t_{n-1}), \dots, X(t_1))$  and  $X(t_n) \mid X(t_{n-1})$ ,  $t_1 < \dots < t_n$ , have the same distribution.*

*Proof* Let  $f(x_1, \dots, x_k)$  denote the density of  $(X(t_1), \dots, X(t_k))$ ,  $k \geq 2$ ,  $t_1 < \dots < t_k$ , and  $f(x_n \mid x_{n-1})$  the density of  $X(t_n) \mid X(t_{n-1})$ . We have



$f(x_n, x_{n-2}, \dots, x_1 \mid x_{n-1}) = f(x_{n-2}, \dots, x_1 \mid x_{n-1})f(x_n \mid x_{n-1})$  since, conditional on present, past and future are independent. This implies  $f(x_n \mid x_{n-1}) = f(x_n, \dots, x_1)/f(x_{n-1}, \dots, x_1) = f(x_n \mid x_{n-1}, \dots, x_1)$  so that  $f(x_n, \dots, x_1) = f(x_n \mid x_{n-1}, \dots, x_1)f(x_{n-1}, \dots, x_1) = f(x_n \mid x_{n-1})f(x_{n-1}, \dots, x_1)$ . Repeated use of the latter formula gives  $f_n(x_1, \dots, x_n) = f(x_1) \prod_{i=2}^n f(x_i \mid x_{i-1})$ , where  $f(x_1)$  denotes the density of  $X(t_1)$ . The latter formula gives  $f_n(x_1, \dots, x_n)/f_{n-1}(x_1, \dots, x_{n-1}) = f(x_n \mid x_{n-1})$ , that is, the second part of the theorem.  $\blacktriangle$

**Theorem 3.30** Let  $f_{X(v)|X(u)}(\cdot \mid \cdot)$ ,  $v > u$ , denote the density of  $X(v) \mid X(u)$ . For  $t_0 < s < t$  the conditional densities of  $X(t) \mid X(t_0)$ ,  $X(t) \mid X(s)$ , and  $X(s) \mid X(t_0)$  are related by

$$f_{X(t)|X(t_0)}(x \mid x_0) = \int_{\mathbb{R}} f_{X(t)|X(s)}(x \mid y) f_{X(s)|X(t_0)}(y \mid x_0) dy, \quad (3.42)$$

referred to as the Chapman–Kolmogorov equation (Exercise 3.14).

**Example 3.46** The time series in Example 3.2 is Markov since  $X_i \mid (X_{i-1}, X_{i-2}, \dots) \stackrel{d}{=} X_i \mid X_{i-1} \sim N(\rho X_{i-1}, 1 - \rho^2)$ . This Markov process has discrete time and continuous state since the index time set  $I$  is countable and the state can take any value on the real line.  $\diamond$

### 3.7.5 Processes with Independent Increments

This class of stochastic processes is essential for our discussion on stochastic differential equations. Considerations in this section are limited to real-valued stochastic processes. Extension to vector-valued processes is direct.

**Definition 3.32** A real-valued process  $X(t)$ ,  $t \geq 0$ , has independent increments if the random variables  $X(t) - X(v)$  and  $X(u) - X(s)$  are independent for all  $s < u \leq v < t$ . A process is said to have stationary independent increments if it has independent increments and the distribution of  $X(t) - X(s)$ ,  $t > s$ , depends only on time lag  $t - s$ , rather than  $s$  and  $t$ .

Brownian motion is an example of a process with stationary independent increments (Definition 3.3). The Poisson and the compound Poisson processes introduced in Example 3.5 also have stationary independent increments.

**Theorem 3.31** The finite dimensional densities of a real-valued process  $X(t)$ ,  $t \geq 0$ , with  $X(0) = 0$  and independent increments are

$$f_n(x_1, \dots, x_n; t_1, \dots, t_n) = f_{Y_1}(x_1) f_{Y_2}(x_2 - x_1) \dots f_{Y_n}(x_n - x_{n-1}), \quad (3.43)$$

where  $n \geq 1$  is an integer,  $0 = t_0 < t_1 < t_2 < \dots < t_n$  denote arbitrary times,  $f_n(x_1, \dots, x_n; t_1, \dots, t_n)$  is the density of  $(X(t_1), \dots, X(t_n))$ , and  $f_{Y_i}$  is the density of  $Y_i = X(t_i) - X(t_{i-1})$ ,  $i = 1, \dots, n$ .

*Proof* Since  $P(X(t_1) \leq x_1, \dots, X(t_n) \leq x_n) = P(Y_1 \leq x_1, Y_1 + Y_2 \leq x_2, \dots, \sum_{k=1}^n Y_k \leq x_n)$ , we have

$$\begin{aligned} P(X(t_1) \leq x_1, \dots, X(t_n) \leq x_n) &= \int_{-\infty}^{x_1} dy_1 f_{Y_1}(y_1) \int_{-\infty}^{x_2 - y_1} dy_2 f_{Y_2}(y_2) \\ &\dots \int_{-\infty}^{x_{n-1} - \sum_{k=1}^{n-2} y_k} dy_{n-1} f_{Y_{n-1}}(y_{n-1}) \int_{-\infty}^{x_n - \sum_{k=1}^{n-1} y_k} dy_n f_{Y_n}(y_n). \end{aligned}$$

The finite dimensional densities of  $X$  result by differentiation of the above equation or from the density of the random vector  $(Y_1, \dots, Y_n)$  and the mapping  $(Y_1, \dots, Y_n) \mapsto (X(t_1), \dots, X(t_n))$ .  $\blacktriangle$

**Theorem 3.32** *If  $X(t)$ ,  $t \geq 0$ , is a real-valued process with stationary independent increments and finite variance, then (1)  $X$  is a Markov process, (2)  $\mu(t) = E[X(t)]$  and  $\sigma(t)^2 = E[(X(t) - \mu(t))^2]$  are linear function of  $t$  provided  $X(0) = 0$ , that is,  $\mu(s+t) = \mu(s) + \mu(t)$  and  $\sigma(s+t)^2 = \sigma(s)^2 + \sigma(t)^2$ , and (3)  $f_{\tau_1+\tau_2} = f_{\tau_2} * f_{\tau_1}$  and  $\varphi_{\tau_1+\tau_2} = \varphi_{\tau_1}\varphi_{\tau_2}$ , where  $\tau_1, \tau_2 > 0$ ,  $f_\tau$  and  $\varphi_\tau$  denote the density and the characteristic functions of  $X(t+\tau) - X(t)$ ,  $\tau > 0$ , and  $*$  denotes convolution.*

*Proof* That  $X$  is Markov follows from (3.43) showing that the conditional random variables  $X(t_n) \mid (X(t_{n-1}), \dots, X(t_1))$  and  $X(t_n) \mid X(t_{n-1})$  have the same density. Since  $X(t+\tau_1+\tau_2) - X(t) = (X(t+\tau_1+\tau_2) - X(t+\tau_1)) + (X(t+\tau_1) - X(t))$  and  $(X(t+\tau_1+\tau_2) - X(t+\tau_1))$  is independent of  $(X(t+\tau_1) - X(t))$ , we have  $f_{\tau_1+\tau_2} = f_{\tau_2} * f_{\tau_1}$ . Similar arguments imply  $\varphi_{\tau_1+\tau_2} = \varphi_{\tau_1}\varphi_{\tau_2}$  and the stated properties of the mean and variance of  $X$ .  $\blacktriangle$

*Example 3.47* The finite dimensional density of order  $n$  of a Brownian motion  $B$  is

$$f_n(x_1, \dots, x_n; t_1, \dots, t_n) = \phi\left(\frac{x_1}{\sqrt{t_1}}\right) \phi\left(\frac{x_2 - x_1}{\sqrt{t_2 - t_1}}\right) \dots \phi\left(\frac{x_n - x_{n-1}}{\sqrt{t_n - t_{n-1}}}\right), \quad (3.44)$$

where  $0 \leq t_1 < \dots < t_n$ . Let  $C(t) = \sum_{k=1}^{N(t)} Y_k$  be a compound Poisson process, where  $\lambda > 0$  is the intensity of the underlying Poisson process  $N$ ,  $\{Y_k\}$  are iid real-valued random variables. The characteristic function of  $(C(t_1), \dots, C(t_n))$  can be obtained simply from the characteristic function of increments  $C(t) - C(s)$ ,  $t \geq s$ , of this process, that is, the function  $E[\exp(iu(C(t) - C(s)))] = \exp(\lambda(t-s)(1 - \varphi_y(u)))$ , where  $\varphi_y(u) = E[\exp(iuY_1)]$ .  $\diamond$

*Proof* The density in (3.44) follows from (3.43) and the independence of increments of  $B$ . The characteristic function of  $C(t) - C(s) \stackrel{d}{=} C(t-s)$ ,  $t > s$ , is

$$\begin{aligned} E[e^{iuC(t-s)}] &= \sum_{n=0}^{\infty} E[e^{iuC(t-s)} \mid N(t-s) = n] P(N(t-s) = n) \\ &= e^{-\lambda(t-s)} \sum_{n=0}^{\infty} (\varphi_y(u))^n \frac{[\lambda(t-s)]^n}{n!} = e^{-\lambda(t-s)[1-\varphi_y(u)]}, \end{aligned}$$

since  $E[\exp(iuC(t-s)) \mid N(t-s) = n] = E[\exp(iu \sum_{k=1}^n Y_k)] = \varphi_Y(u)^n$  and  $P(N(t-s) = n) = [\lambda(t-s)]^n e^{-\lambda(t-s)} / n!$ . The characteristic function of  $(C(t_1), \dots, C(t_n))$  can be obtained from that of  $C(t) - C(s)$  since, for example, the characteristic function of  $(C(t_1), C(t_2), C(t_3))$  has the expression

$$\begin{aligned} & E\{\exp[i(u_1 C(t_1) + u_2 C(t_2) + u_3 C(t_3))]\} \\ &= E\{\exp[i((u_1 + u_2 + u_3)C(t_1) + (u_2 + u_3)(C(t_2) - C(t_1)) \\ &\quad + u_3(C(t_3) - C(t_2)))]\}, \end{aligned}$$

that depends on the characteristic functions of increments of  $C$ .  $\blacktriangle$

### 3.7.6 Continuous Time Martingales

This class of processes is relevant for a broad range of topics considered in the book, including the characterization of the state of physical systems subjected to random noise. Following general considerations on continuous time martingales, we discuss Brownian motion, Poisson, and Lévy processes, and define white noise processes.

The definition and many of the properties of continuous time martingales are similar to those for discrete time martingales. We only provide a brief introduction on continuous time martingales. The reader may consult [3] (Chaps. 2 and 5), [24] (Chap. 2), [6] (Chap. 1), and [25] (Chap. 4) for additional information on this topic.

**Definition 3.33** A real-valued stochastic process  $X$  defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$  is an  $\mathcal{F}_t$ -martingale if (1)  $E[|X(t)|] < \infty$  for all  $t \geq 0$ , (2)  $X$  is  $\mathcal{F}_t$ -adapted, that is,  $X(t) \in \mathcal{F}_t$  for all  $t \geq 0$ , and (3)  $E[X(t) \mid \mathcal{F}_s] = X(s)$  for all  $s \leq t$ . Condition (3) can be replaced by  $E[X(t) \mid \mathcal{F}_s] = X(s \wedge t)$  for all  $t, s \geq 0$ . Submartingales, supermartingales, and  $p$ -integrable martingales are defined as for discrete-time processes (Sect. 2.12).

*Example 3.48* Let  $Y$  be a random variable with finite mean defined on a probability space  $(\Omega, \mathcal{F}, P)$  with a filtration  $(\mathcal{F}_t)_{t \geq 0}$ . The process  $X(t) = E[Y \mid \mathcal{F}_t]$ ,  $t \geq 0$ , is an  $\mathcal{F}_t$ -martingale.  $\diamond$

*Proof* Jensen's inequality for random variables gives  $E[|X(t)|] = E\{|E[Y \mid \mathcal{F}_t]| \} \leq E\{E[|Y| \mid \mathcal{F}_t]\} = E[|Y|]$ , so that  $X(t)$  is integrable.  $X$  is  $\mathcal{F}_t$ -adapted since  $E[Y \mid \mathcal{F}_t]$  is  $\mathcal{F}_t$ -measurable. Properties of the conditional expectation give  $E[X(t) \mid \mathcal{F}_s] = E\{E(Y \mid \mathcal{F}_t) \mid \mathcal{F}_s\} = E[Y \mid \mathcal{F}_s] = X(s)$  for all  $s \leq t$ .  $\blacktriangle$

*Example 3.49* Let  $C(t) = \sum_{k=1}^{N(t)} Y_k$ ,  $t \geq 0$ , be a compound Poisson process, where  $\{Y_k\}$  are iid random variables with mean 0 and  $N(t)$  is a homogeneous Poisson process with intensity  $\lambda > 0$ . Then  $C$  is a martingale with respect to its natural filtration  $\mathcal{F}_t$ .  $\diamond$

*Proof* Since  $E[C(t)] = E\{E[C(t) \mid N(t)]\}$ ,  $E[C(t) \mid N(t)] = 0$ , and  $C(t) \in \mathcal{F}_t$ , the first two defining properties for martingales are satisfied. The third property holds

since, for  $s < t$ ,  $E[C(t) \mid \mathcal{F}_s] = E[(C(t) - C(s)) + C(s) \mid \mathcal{F}_s] = E[C(t) - C(s)] + C(s)$  since  $C(t) - C(s)$  is independent of  $\mathcal{F}_s$  and  $C(s)$  is  $\mathcal{F}_s$ -measurable. This gives  $E[C(t) \mid \mathcal{F}_s] = C(s)$  since  $E[C(t) - C(s)] = 0$  by assumption.  $\blacktriangle$

**Definition 3.34** A real-valued process  $X$  defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$  is an  $\mathcal{F}_t$ -local martingale, submartingale, or supermartingale if there exists an increasing sequence  $T_n$ ,  $n = 1, 2, \dots$ , of  $\mathcal{F}_t$ -stopping times, that is,  $T_n \leq T_{n+1}$ , such that (1)  $\lim_{n \rightarrow \infty} T_n = +\infty$  a.s. and (2) the stopped process  $X^{T_n}(t) = X(t \wedge T_n)$  is an  $\mathcal{F}_t$ -martingale, submartingale, or supermartingale, respectively, for each  $n$ . A sequence  $T_n$  with these properties is called a localizing sequence. Recall that  $T$  is a stopping time if  $\{T \leq t\} \in \mathcal{F}_t$  for all  $t \geq 0$ .

Martingales, submartingales, and supermartingales are local martingales, submartingales, and supermartingales, respectively, since  $T_n = \infty$ ,  $n \geq 1$ , is a localizing sequence for them. We will use local martingales to define stochastic integrals (Sect. 4.4.3).

**Theorem 3.33** *Increments of a martingale  $X$  over non-overlapping intervals are orthogonal provided  $X(t) \in L^2$  for all  $t \geq 0$ .*

*Proof* The expectation  $E[(X(t) - X(s))(X(v) - X(u))]$ ,  $u \leq v \leq s \leq t$ , is equal to  $E\{E[(X(t) - X(s))(X(v) - X(u)) \mid \mathcal{F}_s]\} = E\{(X(v) - X(u))E[X(t) - X(s) \mid \mathcal{F}_s]\}$ , which is 0 since  $X$  is a martingale so that  $E[X(t) - X(s) \mid \mathcal{F}_s] = 0$ .  $\blacktriangle$

**Definition 3.35** Let  $\mathcal{P}$  and  $\mathcal{O}$  be the smallest  $\sigma$ -fields on  $[0, \infty) \times \Omega$  with respect to which all left and right continuous  $\mathcal{F}_t$ -adapted processes are measurable, respectively. A process  $X$  is called predictable if the mapping  $(t, \omega) \mapsto X(t, \omega)$  is  $\mathcal{P}$ -measurable. If this mapping is  $\mathcal{O}$ -measurable, then  $X$  is said to be optional.

Predictable processes allow a peek into the future since they are left continuous. Note also that predictable and optional processes are measurable since the  $\sigma$ -fields  $\mathcal{P}$  and  $\mathcal{O}$  are included in  $\mathcal{B}[0, \infty) \times \mathcal{F}$  and that  $\mathcal{P} \subseteq \mathcal{O}$  ([5], Sect. 3.11.1). The following theorem stated without proof gives an important property of martingales.

**Theorem 3.34** *Every martingale has a unique modification whose samples are right continuous with left limits a.s., that is, every martingale has an optional modification ([6], Corollary 1, p. 8).*

**Theorem 3.35** (Jensen's inequality) *If  $X$  is an  $\mathcal{F}_t$ -martingale and  $\varphi : \mathbb{R} \mapsto \mathbb{R}$  is a convex function such that  $E[|\varphi(X(t))|] < \infty$ ,  $t \geq 0$ , then  $E[\varphi(X(t)) \mid \mathcal{F}_s] \geq \varphi(X(s))$ ,  $s \leq t$ .*

*Proof* Convex functions are continuous so that  $\varphi(X(t))$  is a random variable. Since  $\varphi$  is a convex function, we have  $\varphi(x) = \sup\{l(x) : l(u) \leq \varphi(u), \forall u\}$ , where  $l(u) = au + b$  and  $a, b$  are real constants. Then

$$\begin{aligned} E[\varphi(X(t)) \mid \mathcal{F}_s] &= E[\sup\{l(X(t))\} \mid \mathcal{F}_s] \geq \sup\{E[l(X(t)) \mid \mathcal{F}_s]\} \\ &= \sup\{l(E[X(t) \mid \mathcal{F}_s])\} = \sup\{l(X(s))\} = \varphi(X(s)) \end{aligned}$$

**Table 3.1** Defining properties for Brownian motion, Poisson, and Lévy processes

Brownian motion	Poisson	Lévy
$\mathcal{F}_t$ -adapted starting at zero		
Stationary increments that are independent of the past		
Gaussian increments	Counting process	Continuous in
$B(t) - B(s) \sim N(0, t - s)$	without explosions	probability

by Theorem 2.11, linearity of the expectation operator, and the martingale property of  $X(t)$ .  $\blacktriangle$

The remainder of this section deals with Brownian motion, Poisson, and Lévy processes, use these processes to illustrate properties of continuous time martingale, and develop preliminary tools for solving stochastic equations. Table 3.1 adapted from [5] (Sect. 3.14) summarizes the defining properties of Brownian motion, Poisson, and Lévy processes.

### 3.7.6.1 Brownian Motion

The construction of a Brownian motion process  $B(t)$ ,  $0 \leq t \leq 1$ , may involve the following three steps ([23], Sect. 6.3). First, select a sequence of refining partitions  $p_n = \{k/2^n : k = 0, 1, \dots, 2^n\}$  in  $[0,1]$ , where  $n \geq 1$  is an integer. Second, define a sequence of processes  $B^{(n)}$  in  $[0,1]$  such that (i) the random variables  $B^{(n)}(k/2^n) - B^{(n)}((k-1)/2^n)$  are independent copies of  $N(0, 1/2^n)$  and (ii) the samples of  $B^{(n)}$  are linear in each interval  $[(k-1)/2^n, k/2^n]$ . Third, define  $B(t)$  as the limit of the sequence of processes  $B^{(n)}(t)$  as  $n \rightarrow \infty$ . The limit process  $B(t)$  has continuous samples and stationary independent Gaussian increments.

**Definition 3.36** Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$  be a filtered probability space. A process  $B(t)$ ,  $t \geq 0$ , defined on this space is a Brownian motion if (1) it is  $\mathcal{F}_t$ -adapted and starts at zero, (2) it has stationary increments that are independent of the past, that is,  $B(t) - B(s)$ ,  $t > s$ , is independent of  $\mathcal{F}_s$ , and (3)  $B(t) - B(s) \sim N(0, t - s)$  for  $t > s$  (Table 3.1).

A similar definition holds for  $\mathbb{R}^d$ -valued Brownian motions. In this case, the increments  $B(t) - B(s)$ ,  $t > s$ , are Gaussian vectors with mean zero and covariance matrix  $\gamma(t - s)$ , where  $\gamma$  is an  $(d,d)$ -positive definite matrix. We use for  $B$  a modification of Brownian motion with a.s. continuous samples that is guaranteed to exist ([6], Theorem 26). Note that independence of the past is a stronger requirement than that of independent increments. Let  $X$  be a process adapted to a filtration  $\mathcal{F}_t$ ,  $t \geq 0$ , and take  $0 \leq u < v \leq s < t$ . Since  $X(t) - X(s)$  is independent of  $\mathcal{F}_s$  by hypothesis, it is also independent of  $X(v) - X(u) \in \mathcal{F}_v \subseteq \mathcal{F}_s$ .

**Theorem 3.36** *Brownian motion is measurable, that is, the function  $B : I \times \Omega \rightarrow \mathbb{R}$  is measurable from  $(I \times \Omega, \mathcal{B}(I) \times \mathcal{F})$  to  $(\mathbb{R}, \mathcal{B})$ , where  $I$  is a bounded interval.*

*Proof* Set  $I = [0, 1]$  and let  $B^{(n)}(t, \omega) = B(k2^{-n}, \omega)$ ,  $t \in ((k-1)2^{-n}, k2^{-n}]$ ,  $k = 1, \dots, 2^n$ , with  $B^{(n)}(0, \omega) = 0$ . The process  $B^{(n)}$  has piecewise constant sample paths, depends on a countable number of values of  $B$ , and can be viewed as a random walk with time step  $\Delta t = 2^{-n}$  defined by a sum of independent Gaussian variables with mean zero and variance  $2^{-n}$  for each  $n$ , so that the function  $(t, \omega) \mapsto B^{(n)}(t, \omega)$  is measurable from  $\mathcal{B}(I) \times \mathcal{F}$  to  $(\mathbb{R}, \mathcal{B})$ . The measurable mapping  $(t, \omega) \mapsto B^{(n)}(t, \omega)$  converges to  $(t, \omega) \mapsto B(t, \omega)$  pointwise by the continuity of the sample paths of  $B$  so that the function  $B(t, \omega) = \lim_{n \rightarrow \infty} B^{(n)}(t, \omega)$  is measurable, since limits of measurable functions are measurable.  $\blacktriangle$

*Example 3.50* Let  $p_n = (0 = t_0 \leq t_1 \leq \dots \leq t_n = t)$  be a sequence of partitions of  $[0, t]$  such that its mesh  $\Delta(p_n) = \max_{1 \leq k \leq n} (t_k - t_{k-1}) \rightarrow 0$  as  $n \rightarrow \infty$ . Then  $E[\sum_{k=1}^n (\Delta B_k)^2] = t$  for each  $n$  and  $\sum_{k=1}^n (\Delta B_k)^2 \rightarrow t$  in m.s. as  $n \rightarrow \infty$ , where  $\Delta B_k = B(t_k) - B(t_{k-1})$ .  $\diamond$

*Proof* Since  $\Delta B_k \sim N(0, t_k - t_{k-1})$  are independent random variables, we have  $E[\sum_{k=1}^n (\Delta B_k)^2] = \sum_{k=1}^n E[(\Delta B_k)^2] = \sum_{k=1}^n (t_k - t_{k-1}) = t$ . For  $\Delta t_k = t_k - t_{k-1}$  we have

$$\begin{aligned} E\left[\left(\sum_{k=1}^n (\Delta B_k)^2 - t\right)^2\right] &= \sum_{k,l=1}^n E[(\Delta B_k)^2 (\Delta B_l)^2] - 2t \sum_{k=1}^n E[(\Delta B_k)^2] + t^2 \\ &= 3 \sum_{k=1}^n (\Delta t_k)^2 + \sum_{k,l=1, k \neq l}^n \Delta t_k \Delta t_l - t^2 = 2 \sum_{k=1}^n (\Delta t_k)^2, \end{aligned}$$

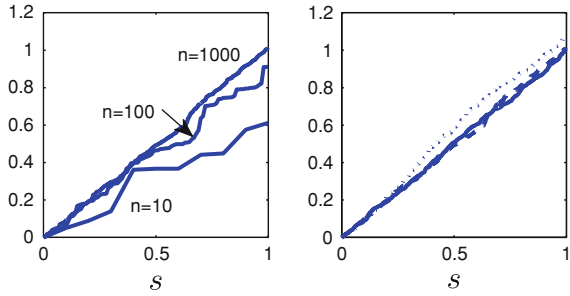
and  $\sum_{k=1}^n (\Delta t_k)^2 \leq (\max_k \Delta t_k) \sum_{k=1}^n \Delta t_k = (\max_k \Delta t_k) t$ , so that  $E[(\sum_{k=1}^n (\Delta B_k)^2 - t)^2] \rightarrow 0$  as  $n \rightarrow \infty$ .  $\blacktriangle$

**Theorem 3.37** *If  $p_n$  is a sequence of refining partitions of  $[0, t]$  such that  $\Delta p_n \rightarrow 0$  as  $n \rightarrow \infty$ , then  $\lim_{n \rightarrow \infty} \sum_{k=1}^n (\Delta B_k)^2 = t$  a.s., where  $\Delta B_k = B(t_k) - B(t_{k-1})$  ([6], Theorem 28, p. 18).*

The process  $[B](s) = \sum_{k=1}^n (B(t_k \wedge s) - B(t_{k-1} \wedge s))^2$ ,  $s \in [0, t]$ , defined by square of increments of Brownian motion is referred to as the quadratic variation of  $B$ , and can be defined for more general processes as we will see in a subsequent chapter. Theorem 3.37 justifies the notation  $(dB(t))^2 = dt$  used frequently in applications. The left panel of Fig. 3.6 shows three samples of  $[B](s)$  for  $s \in [0, t]$ ,  $t = 1$ , and  $t_k = k/n$ ,  $k = 0, 1, \dots, n$ , corresponding to  $n=10, 100$ , and  $1000$ , that is, refining partitions. The right panel in the figure shows three samples of  $[B](s)$  for  $n = 1000$ . If  $n$  is small, the samples of  $[B](s)$  differ from each other and from the identity function  $t \mapsto t$ . These samples nearly coincide with  $t \mapsto t$  for  $n = 1000$ .

**Theorem 3.38** *The samples of a Brownian motion process  $B$  are of unbounded variation a.s. in any bounded interval.*

**Fig. 3.6** Samples of  $[B](s) = \sum_{k=1}^n (B(t_k \wedge s) - B(t_{k-1} \wedge s))^2$ ,  $s \in [0, 1]$ , for  $t_k = k/n$



*Proof* Let  $p_n = (0 = t_0 \leq t_1 \leq \dots \leq t_n = t)$  be a sequence of refining partitions of  $[0, t]$  such that  $\Delta(p_n) \rightarrow 0$  as  $n \rightarrow \infty$ . The left side of the inequality

$$\sum_{k=1}^n [B(t_k) - B(t_{k-1})]^2 \leq \max_k |B(t_k) - B(t_{k-1})| \sum_{k=1}^n |B(t_k) - B(t_{k-1})|$$

converges a.s. to  $t$  as  $n \rightarrow \infty$  by Theorem 3.37 and  $\max_k |B(t_k) - B(t_{k-1})|$  converges a.s. to 0 as  $n \rightarrow \infty$  by the continuity of the samples of  $B$ . To satisfy the above inequality,  $\sum_{k=1}^n |B(t_k) - B(t_{k-1})|$  must approach infinity a.s. as  $n \rightarrow \infty$  (see also [6], Theorem 29). ▲

### 3.7.6.2 Poisson and Compound Poisson Processes

Let  $\{T_n, n = 0, 1, 2, \dots, \infty\}$  with  $T_0 = 0$  be a strictly increasing sequence of positive random variables, and define the counting process

$$N(t) = \sum_{n=1}^{\infty} 1(t \geq T_n), \quad t \geq 0. \quad (3.45)$$

We say that  $N(t)$  is without explosion if  $\sup_n T_n = \infty$ , so that, it is finite a.s. in any bounded time interval.

**Definition 3.37** A process  $N(t)$ ,  $t \geq 0$ , defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$  is said to be a Poisson process if it (1) is  $\mathcal{F}_t$ -adapted and starts at zero, (2) has stationary increments that are independent of the past, that is,  $N(t) - N(s)$ ,  $t > s$ , has the same distribution as  $N(t - s)$  and is independent of  $\mathcal{F}_s$ , and (3) is counting without explosion (Table 3.1).

The defining properties of Poisson processes imply that the jump times  $\{T_n\}$  must be  $\mathcal{F}_t$ -stopping times, since  $N$  is  $\mathcal{F}_t$ -adapted ([6], Theorem 22, p. 14), and that  $N(t)$  is a Poisson random variable with distribution

$$P(N(t) = n) = \frac{\lambda t}{n!} e^{-\lambda t}, \quad n = 0, 1, \dots, \quad (3.46)$$

where  $\lambda > 0$  denotes an intensity parameter ([6], Theorem 23, p. 14).

The mean and covariance functions of a Poisson process  $N$  with intensity  $\lambda > 0$  are  $E[N(t)] = \lambda t$  and  $\text{Cov}[N(s)N(t)] = \lambda(s \wedge t)$ . The characteristic function of  $N(t)$  is  $\varphi(u; t) = \exp[-\lambda t(1 - e^{iu})]$  (Exercise 3.16).

*Example 3.51* Let  $N(t)$  be a Poisson process with intensity  $\lambda > 0$ . Then  $N(t)$  is a submartingale and its compensated version,  $N(t) - \lambda t$ , is a martingale. The process  $(N(t) - \lambda t)^2 - \lambda t$  is also a martingale.  $\diamond$

*Proof*  $N(t)$  and  $N(t) - \lambda t$  have finite mean and are  $\mathcal{F}_t$ -adapted. For  $t > s$ ,  $E[N(t) | \mathcal{F}_s] = E[(N(t) - N(s)) + N(s) | \mathcal{F}_s] = E[N(t) - N(s)] + N(s) = \lambda(t - s) + N(s) \geq N(s)$  since  $N(t) - N(s)$  is independent of  $\mathcal{F}_s$  and  $N(s)$  is  $\mathcal{F}_s$ -measurable. Hence,  $N$  is a submartingale and the compensate Poisson process is a martingale.

For  $(N(t) - \lambda t)^2 - \lambda t$ , note that  $E[(N(t) - \lambda t)^2] - \lambda t < \infty$  and  $(N(t) - \lambda t)^2 - \lambda t$  is  $\mathcal{F}_t$ -adapted. For  $t \geq s$ ,

$$\begin{aligned} E[(N(t) - \lambda t)^2 | \mathcal{F}_s] &= E[(N(t) - \lambda t) - (N(s) - \lambda s) + (N(s) - \lambda s)]^2 | \mathcal{F}_s] \\ &= E[(N(t) - N(s) - \lambda(t - s))^2 | \mathcal{F}_s] + E[(N(s) - \lambda s)^2 | \mathcal{F}_s] \\ &\quad + 2E[(N(t) - N(s) - \lambda(t - s))(N(s) - \lambda s) | \mathcal{F}_s] = \lambda(t - s) + (N(s) - \lambda s)^2 \end{aligned}$$

since  $N(t) - N(s)$  is independent of  $\mathcal{F}_s$ ,  $N(s) \in \mathcal{F}_s$ , and  $N(t) - N(s) - \lambda(t - s)$  is a martingale. Hence,  $(N(t) - \lambda t)^2 - \lambda t$  is a martingale.  $\blacktriangle$

*Example 3.52* Set  $M(t) = N(t) - t$ ,  $t \geq 0$ , where  $N$  is a Poisson process with unit intensity. The quadratic variation process of  $M(t) = N(t) - t$ ,  $t \geq 0$ , is  $[M](t) = N(t)$ , where  $[M]$  is defined in the same manner as  $[B]$  shown in Fig. 3.6.  $\diamond$

*Proof* Let  $p_n = \{kt/n, k = 0, 1, \dots, n\}$  be a partition of  $[0, t]$  and set  $\Delta M_k = M(kt/n) - M((k-1)t/n)$ . If  $n$  is sufficiently large, the intervals  $((k-1)t/n, kt/n]$  contain at most a jump of  $N$  so that  $\Delta M_k = -1/n$  if there is no jump and  $\Delta M_k = 1 - 1/n$  if there is a jump. Let  $J_n(t)$  be the collection of intervals  $((k-1)t/n, kt/n]$  containing a jump of  $N$ , so that for sufficiently large  $n$  we have

$$\begin{aligned} S_n(M, M) &= \sum_{k \notin J_n(t)} (\Delta M_k)^2 + \sum_{k \in J_n(t)} (\Delta M_k)^2 = \sum_{k \notin J_n(t)} 1/n^2 + \sum_{k \in J_n(t)} (1 - 1/n)^2 \\ &= (n - N(t))/n^2 + N(t)(1 - 1/n)^2 \rightarrow N(t), \quad \text{as } n \rightarrow \infty, \end{aligned}$$

that is, the sum of the square of increments of  $M$  in  $[0, t]$  over partitions of this time interval converges a.s. to  $N(t)$  as partition mesh decreases to 0.  $\blacktriangle$

*Example 3.53* Let  $B$  be a Brownian motion and  $N$  denote a Poisson process with intensity  $\lambda > 0$ . If  $B$  and  $N$  are independent of each other, the quadratic variation process of  $B + \tilde{N}$  is  $[B + \tilde{N}](t) = t + N(t)$ , where  $\tilde{N}(t) = N(t) - \lambda t$ .  $\diamond$

*Proof* Let  $p_n = \{kt/n, k = 0, 1, \dots, n\}$  be a sequence of partitions of a time interval  $[0, t]$  and consider the sums



$$\begin{aligned}
S_n(B + \tilde{N}, B + \tilde{N}) &= \sum_k (\Delta B_k + \Delta \tilde{N}_k)^2 = \sum_k (\Delta B_k)^2 + \sum_k (\Delta \tilde{N}_k)^2 \\
&\quad + 2 \sum_k (\Delta B_k)(\Delta \tilde{N}_k),
\end{aligned}$$

where  $\Delta B_k = B(t_k) - B(t_{k-1})$ ,  $\Delta \tilde{N}_k = \tilde{N}(t_k) - \tilde{N}(t_{k-1})$ , and  $t_k = kt/n$ . If the sequence of partitions  $p_n$  is refining and  $\Delta(p_n) \rightarrow 0$  as  $n \rightarrow \infty$ , then  $\sum_k (\Delta B_k)^2$  and  $\sum_k (\Delta \tilde{N}_k)^2$  converge to  $t$  and  $N(t)$  a.s., while  $\sum_k (\Delta B_k)(\Delta \tilde{N}_k)$  converges to zero since the Brownian motion has continuous samples so that  $|\sum_k (\Delta B_k)(\Delta \tilde{N}_k)| \leq \sum_k |\Delta B_k| |\Delta \tilde{N}_k| \leq \max_k |\Delta B_k| \sum_k |\Delta \tilde{N}_k|$  and  $\max_k |\Delta B_k| \rightarrow 0$  as  $n \rightarrow \infty$ . Hence,  $S_n(B + \tilde{N}, B + \tilde{N})$  converges a.s. to  $t + N(t)$  as  $n \rightarrow \infty$ . If the sequence of partitions  $p_n$  is not refining, the convergence to  $t + N(t)$  as  $n \rightarrow \infty$  is in m.s. and probability.  $\blacktriangle$

Compound Poisson processes have been introduced in Example 3.5 as sums of iid random variables arriving at the jump times of Poisson processes. This definition can also be given in the form

$$C(t) = \sum_{n=1}^{\infty} Y_n 1(t \geq T_n) = \sum_{n=1}^{N(t)} Y_n, \quad t \geq 0, \quad (3.47)$$

where  $N$  is a Poisson process with intensity  $\lambda > 0$ , the random variables  $\{T_n\}$  denote the jump times of  $N$ , and  $\{Y_n\}$  are iid random variables.

If the random variables  $\{Y_n\}$  in (3.47) have finite variance, then  $E[C(t)] = \lambda t E[Y_1]$  and  $\text{Cov}[C(s), C(t)] = \lambda(s \wedge t) E[Y_1^2]$  are the mean and the covariance functions of  $C(t)$  (Example 3.16). The characteristic function of  $C(t)$  is  $\varphi(u; t) = E[\exp(iu C(t))] = \exp[-\lambda t(1 - \varphi_{Y_1}(u))]$  since

$$\begin{aligned}
E[\exp(iu C(t))] &= E\left\{E\left[\exp\left(iu \sum_{k=1}^{N(t)} Y_k\right) \mid N(t)\right]\right\} = E\{\varphi_{Y_1}(u)^{N(t)}\} \\
&= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \varphi_{Y_1}(u)^n = \exp[-\lambda t(1 - \varphi_{Y_1}(u))],
\end{aligned}$$

where  $\varphi_{Y_1}$  denotes the characteristic function of  $Y_1$ .

*Example 3.54* Let  $C$  be a compound Poisson process defined by (3.47). The quadratic variation of the compensated compound Poisson process  $\tilde{C}(t) = C(t) - \lambda t E[Y_1]$  is  $[\tilde{C}](t) = \sum_{k=1}^{N(t)} Y_k^2$ .  $\diamond$

*Proof* Let  $p_n = \{k/n, k = 0, 1, \dots, n\}$  be a sequence of refining partitions for a time interval  $[0, t]$ , and consider a sufficiently large  $n$  such that  $C$  can have at most a jump in the intervals  $((k-1)t/n, kt/n]$  of this partition. The increments  $\Delta \tilde{C}_k = \tilde{C}(kt/n) - \tilde{C}((k-1)t/n)$  of  $\tilde{C}$  are equal to  $-\lambda E[Y_1]/n + Y_k$  and  $-\lambda E[Y_1]/n$

if there is and there is no jump of  $N(t)$  in  $((k-1)t/n, kt/n]$ , respectively. Let  $J_n(t)$  be the collection of intervals  $((k-1)t/n, kt/n]$  in which  $N$  has a jump. We have

$$\begin{aligned} \sum_{k=1}^n (\Delta \tilde{C}_k)^2 &= \sum_{k \notin J_n(t)} (\Delta \tilde{C}_k)^2 + \sum_{k \in J_n(t)} (\Delta \tilde{C}_k)^2 \\ &= (n - N(t))(\lambda t E[Y_1]/n)^2 + \sum_{k=1}^{N(t)} (-\lambda t E[Y_1]/n + Y_k)^2 \end{aligned}$$

which converges to  $\sum_{k=1}^{N(t)} Y_k^2$  as  $n \rightarrow \infty$ . Note also that the quadratic variation of  $C(t)$  is  $[C](t) = \sum_{k=1}^{N(t)} Y_k^2$ .  $\blacktriangle$

We now give an alternative definition of compound Poisson processes that uses a random measure specifying the number of jumps of  $C$  in rectangles with sides time and space. The definition provides a link between compound Poisson processes and Lévy processes, that are discussed in the following section.

**Definition 3.38** Let  $\mathcal{M}(t, dy)$  be a random measure giving the number of jumps of  $C$  in the rectangle  $(0, t] \times (y, y + dy]$  with expectation  $E[\mathcal{M}(t, dy)] = \lambda t dF(y) = \mu(dy)t$ , where  $\mu(dy) = \lambda dF(y)$  and  $F$  denotes the distribution of  $Y_1$ . Then

$$C(t) = \int_{\mathbb{R}} y \mathcal{M}(t, dy), \quad (3.48)$$

is a compound Poisson process ([26], Theorem 3.3.2).

**Theorem 3.39** Let  $C$  be a compound Poisson process and  $\Lambda$  be a Borel set in  $\mathbb{R}$ . Then

$$C^\Lambda(t) = \sum_{k=1}^{N(t)} Y_k 1(Y_k \in \Lambda) = \sum_{0 < s \leq t} \Delta C(s) 1(\Delta C(s) \in \Lambda) = \int_{\Lambda} y \mathcal{M}(t, dy) \quad (3.49)$$

is a compound Poisson process corresponding to a Poisson process with intensity  $\tilde{\lambda} = \lambda P(Y_1 \in \Lambda)$ , where  $\Delta C(s) = C(s) - C(s-)$  and  $C(s-) = \lim_{s \uparrow t} C(s)$ .

*Proof* Since  $C^\Lambda$  retains the jumps of  $C$  in  $\Lambda$ , it is said to be a thinned version of  $C$ . Let  $\tilde{Y}_k = Y_k 1(Y_k \in \Lambda)$  be the jumps of  $C$  taking values in  $\Lambda$ . The characteristic function of  $\tilde{Y}_1$  is

$$\begin{aligned} \varphi_{\tilde{Y}_1}(u) &= E[e^{iuY_1 1(Y_1 \in \Lambda)}] = \int_{\Lambda} e^{iu y 1(y \in \Lambda)} dF_{Y_1}(y) + \int_{\mathbb{R} \setminus \Lambda} e^{iu y 1(y \in \Lambda)} dF_{Y_1}(y) \\ &= \int_{\Lambda} e^{iu y} dF_{Y_1}(y) + \int_{\mathbb{R} \setminus \Lambda} dF_{Y_1}(y) = \varphi_{Y_1}^{(\Lambda)}(u) + P(Y_1 \notin \Lambda) \end{aligned}$$

so that  $\lambda t(1 - \varphi_{\tilde{Y}_1}(u)) = \lambda t(P(Y_1 \in \Lambda) - \varphi_{Y_1}^{(\Lambda)}(u)) = \tilde{\lambda} t(1 - \tilde{\varphi}(u))$ , where  $\tilde{\lambda} = \lambda P(Y_1 \in \Lambda)$ ,  $\varphi_{Y_1}^{(\Lambda)}(u) = \int_{\Lambda} e^{iu y} dF_{Y_1}(y)$ , and  $\tilde{\varphi}(u) = \varphi_{Y_1}^{(\Lambda)}(u)/P(Y_1 \in \Lambda)$  is the

characteristic function of  $Y_1 \mid (Y_1 \in \Lambda)$ . Since  $\varphi_{C^\Lambda(t)}(u) = \exp[\tilde{\lambda}t(1 - \tilde{\varphi}(u))]$ ,  $C^\Lambda$  is a compound Poisson process with the stated properties.  $\blacktriangle$

### 3.7.6.3 Lévy Processes

We define Lévy processes, review some of their properties, and present exact and approximate representations for these processes. An extensive discussion on Lévy processes can be found in [1, 3].

**Definition 3.39** Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$  be a filtered probability space. A real-valued process  $X(t)$ ,  $t \geq 0$ , defined on this space is a Lévy process if it (1) is  $\mathcal{F}_t$ -adapted and starts at zero, (2) has stationary increments that are independent of the past, that is,  $X(t) - X(s)$ ,  $t > s$ , has the same distribution as  $X(t - s)$  and is independent of  $\mathcal{F}_s$ , and (3) is continuous in probability (Table 3.1).

*Example 3.55* It can be shown that the characteristic function of a Lévy process  $X$  has the functional form

$$\varphi(u; t) = E[e^{iuX(t)}] = e^{-t\psi(u)}, \quad t \geq 0, \quad u \in \mathbb{R}, \quad (3.50)$$

where  $\psi$  is a continuous function with  $\psi(0) = 0$  ([6], Sect. 4). The Lévy process corresponding to  $\varphi(u; t)$  with  $\psi(u) = |u|^\alpha$ ,  $\alpha \in (0, 2]$ , may or may not be an  $\mathcal{F}_t$ -martingale depending on the value of  $\alpha$ .  $\diamond$

*Proof* Let  $f$  denote the density of  $X(t) - X(s)$ ,  $t > s$ . For every  $\varepsilon > 0$  the probability  $P(|X(t) - X(s)| > \varepsilon) = \int_{(-\varepsilon, \varepsilon)^c} f(x) dx$  converges to zero as  $(t - s) \rightarrow 0$  since  $f(x) = 1/(2\pi) \int_{-\infty}^{\infty} e^{iux} e^{-(t-s)|u|^\alpha} du \rightarrow \delta(0)$  as  $(t - s) \rightarrow 0$ .

It can be shown that  $E[|X(t)|^p]$  is finite if and only if  $p \in (0, \alpha)$  ([1], Proposition 1.2.16). Hence,  $X$  is not a martingale for  $\alpha \leq 1$  because the expectation  $E[|X(t)|]$  is not bounded but this process is an  $\mathcal{F}_t$ -martingale for  $\alpha > 1$ . If  $\alpha = 2$ , the characteristic function of  $X(t)$  is  $\varphi(u; t) = e^{-tu^2}$  so that  $X(t) \sim N(0, 2t)$  is a square integrable martingale that has the same distribution as  $\sqrt{2}B(t)$ , where  $B$  denotes a Brownian motion process.  $\blacktriangle$

The following four theorems gives essential properties of Lévy processes. Note that these properties are similar to those of compound Poisson processes.

**Theorem 3.40** *Sums of independent Lévy processes are Lévy processes.*

*Proof* Let  $X_1$  and  $X_2$  be independent Lévy processes defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ . Then  $X = X_1 + X_2$  is  $\mathcal{F}_t$ -adapted and has stationary increments that are independent of the past as the sum of two processes with these properties. It remains to show that  $X$  is continuous in probability. Set  $A_i = \{|X_i(t) - X_i(s)| > \varepsilon/2\}$ ,  $i = 1, 2$ , and  $A = \{|X(t) - X(s)| > \varepsilon\}$  for some  $\varepsilon > 0$ . Since  $A^c \supseteq A_1^c \cap A_2^c$ , we have  $A \subseteq A_1 \cup A_2$  so that  $P(A) \leq P(A_1) + P(A_2)$  implying  $P(|X(t) - X(s)| > \varepsilon) \rightarrow 0$  as  $|t - s| \rightarrow 0$  by the continuity of  $X_1$  and  $X_2$ .  $\blacktriangle$

**Theorem 3.41** *Lévy processes have unique modifications that are Lévy with right continuous samples that have left limits ([6], Theorem 30, p. 21), preserve their properties if restarted at stopping times ([6], Theorem 32, p. 23), and have only jump discontinuities ([6], p. 6).*

**Theorem 3.42** *Let  $X$  be a Lévy process and  $\Lambda$  denote a Borel set such that  $0 \notin \overline{\Lambda}$ . The associated jump process  $J^\Lambda(t) = \sum_{0 < s \leq t} \Delta X(s) 1(\Delta X(s) \in \Lambda) = \int_\Lambda y \mathcal{M}(t, dy)$  and  $X - J^\Lambda$  are Lévy processes, where  $\mathcal{M}$  is a random measure defining the Poisson process  $J^\Lambda(t)$  ([6], Theorem 37, p. 27).*

This theorem implies that the process

$$Z^a(t) = X(t) - \sum_{0 < s \leq t} \Delta X(s) 1(|\Delta X(s)| > a) = X(t) - \int_{|x| > a} y \mathcal{M}(t, dy) \quad (3.51)$$

consisting of  $X$  from which jumps of magnitude larger than  $a > 0$  have been removed is also Lévy. Moreover, the moments of any order of  $Z^a$  are bounded ([6], Theorem 34, p. 25).

**Theorem 3.43** *The jump processes  $J^{\Lambda_i}$  in Theorem 3.42 corresponding to Borel sets  $\Lambda_1$  and  $\Lambda_2$  such that  $0 \notin \overline{\Lambda_i}$ ,  $i = 1, 2$ , and  $\Lambda_1 \cap \Lambda_2 = \emptyset$  are independent Lévy processes ([6], Theorem 39, p. 30).*

Following are exact and approximate representations for Lévy processes. The exact representation is given by the Lévy decomposition theorem and the Lévy-Khintchine formula, and is primarily useful for theoretical developments ([3], Chap. 2, [6] Sect. I.4, and [1]). The approximate representation is based on features of the large and small jumps of Lévy processes, and is most useful in applications. Both representations involve the Lévy measure  $\lambda_L(\Lambda) = E[N^\Lambda(1)]$ , that is, the intensity of the Poisson process  $N^\Lambda(t) = \sum_{0 < s \leq t} 1(\Delta X(s) \in \Lambda)$ , where  $\Lambda \in \mathcal{B}$  is a Borel set such that  $0 \notin \overline{\Lambda}$ ,  $\overline{\Lambda}$  denotes the closure of  $\Lambda$ , and  $\Delta X(s) = X(s) - X(s-)$ .

**Theorem 3.44** (Lévy decomposition) *If  $X$  is a Lévy process, then it admits the decomposition*

$$X(t) = B(t) + \int_{|y| < 1} y(\mathcal{M}(t, dy) - t\lambda_L(dy)) + \beta t + \sum_{0 < s \leq t} \Delta X(s) 1(|\Delta X(s)| \geq 1), \quad (3.52)$$

where  $B$  is a Brownian motion,  $\mathcal{M}$  is a random measure,  $\lambda_L$  is a measure on  $\mathbb{R} \setminus \{0\}$  such that  $\int \min(1, y^2) \lambda_L(dy) < \infty$ ,  $N^\Lambda(t) = \int_\Lambda \mathcal{M}(t, dy)$  is a Poisson process with intensity  $\lambda_L(\Lambda)$  for any Borel set  $\Lambda$ ,  $0 \notin \overline{\Lambda}$  such that, if  $\Gamma$  is another Borel set with the properties  $0 \notin \overline{\Gamma}$  and  $\Gamma \cap \Lambda = \emptyset$ , then  $N^\Lambda$  and  $N^\Gamma$  are independent,  $N^\Lambda$  is independent of  $B$ , and  $\beta = E[X(1) - J^\Lambda(1)] \in \mathbb{R}$  ([6], Theorem 42, p. 32).

**Theorem 3.45** (Lévy-Khintchine formula) *A Lévy process  $X$  is defined by its characteristic function that has the expression  $\varphi(u; t) = E[e^{iuX(t)}] = e^{-t\psi(u)}$ , where*

$$\psi(u) = \frac{\sigma^2}{2}u^2 + \int_{|y|<1} (1 - e^{iuy} + iuy)\lambda_L(dy) - i\beta u + \int_{|y|\geq 1} (1 - e^{iuy})\lambda_L(dy), \quad (3.53)$$

and  $\lambda_L$ ,  $\sigma^2$ , and  $\beta$  need to be specified. The parameters  $\lambda_L$ ,  $\sigma^2$ , and  $\beta$  define a Lévy process uniquely in distribution ([6], Theorem 43, p. 32).

The approximate representation for a Lévy process  $X(t)$  has two components. The first is the process  $\sum_{0 < s \leq t} \Delta X(s) 1(|\Delta X(s)| > a)$  defined by the jumps of  $X(t)$  with magnitude exceeding  $a > 0$ , and is represented exactly. The second is the process  $Z^a(t)$  in (3.51), and is described approximately by a scaled Brownian motion with scaling factor depending on  $a$ .

*Example 3.56* Let  $X(t)$ ,  $t \geq 0$ , be a Lévy process with characteristic function  $\varphi(u; t) = E[e^{iuX(t)}] = \exp(-t|u|^\alpha)$ , where  $\alpha \in (0, 2]$ . Figure 3.7 shows a sample of  $X$  and a sample of a Brownian motion  $B$  and the corresponding samples of the quadratic variation processes  $[B]$  and  $[X]$ . The sample of  $X$  exhibits jump discontinuities which are emphasized in the sample of  $[X]$ . The steady increase of  $[X]$  resembling  $[B]$  is marked by jumps that are typical to the quadratic variation of a compound Poisson process.  $\diamond$

Let  $X = L_\alpha$ ,  $0 < \alpha \leq 2$ , be a symmetric  $\alpha$ -stable process, that is, a Lévy process with characteristic function

$$\varphi(u; t) = E[e^{iuL_\alpha(t)}] = \exp(-t|u|^\alpha), \quad t \geq 0, \quad u \in \mathbb{R}, \quad (3.54)$$

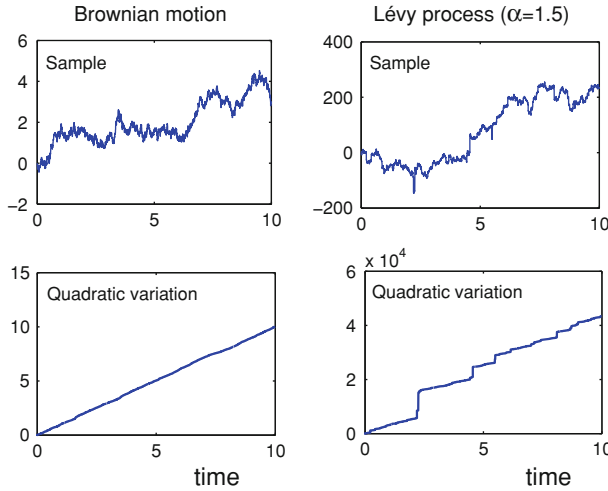
given by (3.53) with  $\sigma = 0$ ,  $\beta = 0$ , and the Lévy measure

$$\lambda_L(dy) = \frac{\alpha c_\alpha}{2} |y|^{-(\alpha+1)} dy, \quad y \in \mathbb{R} \setminus \{0\}, \quad (3.55)$$

where  $c_\alpha = (1 - \alpha)/[\Gamma(2 - \alpha) \cos(\pi\alpha/2)]$  for  $\alpha \neq 1$  and  $c_\alpha = 2/\pi$  for  $\alpha = 1$  ([1, Property 1.2.15]). The definition in (3.54) implies that the increments  $dL_\alpha(t)$  of  $L_\alpha$  are symmetric  $\alpha$ -stable random variables with scale  $(dt)^{1/\alpha}$  centered at 0.

There are notable similarities and differences between compound Poisson and  $\alpha$ -stable processes. Let  $\mathcal{M}(dt, dy)$  be a random measure defining the number of jumps in the infinitesimal rectangle  $(t, t + \Delta t] \times (y, y + dy]$ . The expectation of  $\mathcal{M}(dt, dy)$  is  $\lambda dt dF(y)$  for the compound Poisson process  $C$  in (3.48) and  $\lambda_L(dy) dt$  for an  $\alpha$ -stable process  $L_\alpha$ , where  $F$  denotes the distribution of the jumps  $\{Y_k\}$  of  $C$  and  $\lambda_L$  is the Lévy measure. Let  $\Lambda$  be an arbitrary Borel set in  $\mathbb{R}$ . The integral  $(1/dt) \int_\Lambda \mathcal{M}(dt, dy) dy$  gives the average number of jumps in  $\Lambda$  per unit of time. This number is  $\lambda \int_\Lambda dF(y) \leq \lambda < \infty$  for  $C$  but can be finite or not for  $L_\alpha$  depending on  $\Lambda$ , for example,

$$\begin{aligned} (1/dt) \int_{(-a, a)^c} \mathcal{M}(dt, dy) dy &= 2 \int_a^\infty \lambda_L(dy) = \frac{c_\alpha}{a^\alpha} < \infty \quad \text{and} \\ (1/dt) \int_{(-a, a)} \mathcal{M}(dt, dy) dy &= 2 \int_{0+}^a \lambda_L(dy) = \infty, \end{aligned} \quad (3.56)$$



**Fig.3.7** Samples of a Brownian motion and a Lévy process and of corresponding quadratic variations

that is, the average number of jumps of  $L_\alpha$  with magnitude larger and smaller than  $a > 0$  is finite and unbounded, respectively.

Consider the process  $\sum_{0 \leq s \leq t} \Delta L_\alpha(s) 1(|\Delta L_\alpha(s)| > a)$  in (3.51), that is, the compound Poisson process

$$C_{\alpha,a}(t) = \sum_{k=1}^{N_a(t)} Y_{a,k}, \quad t \geq 0, \quad (3.57)$$

where  $\{Y_{a,k}\}$  denote the jumps of  $L_\alpha$  with magnitude larger than  $a > 0$  and  $N_a$  is a Poisson counting process with intensity  $\lambda_a = c_\alpha/a^\alpha$ . Recall that  $\{Y_{a,k}\}$  are independent identically distributed random variables. The distribution of  $Y_{a,1}$  with  $\lambda_L$  in (3.55) is

$$\begin{aligned} F_a(y) &= \frac{\int_{(-a,a)^c \cap (-\infty, y)} \lambda_L(dz)}{\int_{(-a,a)^c} \lambda_L(dz)} = \frac{\alpha a^\alpha}{2} \int_{(-a,a)^c \cap (-\infty, y)} z^{-(\alpha+1)} dz \\ &= \frac{\alpha a^\alpha}{2} \left[ \frac{1}{\alpha|y|^\alpha} 1(y < -a) + \frac{1}{\alpha a^\alpha} 1(y \geq -a) + \frac{1 - (a/y)^\alpha}{\alpha a^\alpha} 1(y \geq a) \right], \quad y \in \mathbb{R}, \end{aligned} \quad (3.58)$$

so that the density and the characteristic functions of  $Y_{a,1}$  are

$$f_a(y) = \frac{\alpha a^\alpha}{2} \left[ |y|^{-(\alpha+1)} 1(y < -a) + y^{-(\alpha+1)} 1(y \geq a) \right], \quad y \in \mathbb{R}, \quad (3.59)$$

and

$$\varphi_a(u) = \alpha a^\alpha \int_a^\infty \cos(uy) y^{-(\alpha+1)} dy, \quad u \in \mathbb{R}, \quad (3.60)$$

respectively. The asymptotic behavior of the distribution and density functions in (3.58) and (3.59) as  $y \rightarrow \infty$  is  $y^{-\alpha}$  and  $y^{-(\alpha+1)}$ , respectively.

Let  $L_{\alpha,a}$  be a stochastic process obtained from  $L_\alpha$  by excluding the jumps of  $L_\alpha$  with magnitude larger than  $a > 0$ , that is, the process

$$L_{\alpha,a}(t) = L_\alpha(t) - C_{\alpha,a}(t) \quad (3.61)$$

with  $C_{\alpha,a}$  in (3.57). Note that (1)  $L_{\alpha,a}$  and  $C_{\alpha,a}$  are independent Lévy processes ([6], Theorem 39, p. 30), (2)  $L_{\alpha,a}$  has finite absolute moments of any order since it is a Lévy process with bounded jumps ([6], Theorem 34, p. 25), (3)  $L_{\alpha,a}$  has mean 0 since its density is an even function, (4) the variance of  $L_{\alpha,a}(t)$  is linear in  $t$  since  $L_{\alpha,a}(t)$  has stationary independent increments, and (5) the approximation

$$L_\alpha(t) \simeq \tilde{L}_{\alpha,a}(t) = \sigma(\alpha, a)B(t) + C_{\alpha,a}(t), \quad t \geq 0, \quad (3.62)$$

holds in the sense that  $\sigma(\alpha, a)^{-1}(L_\alpha(t) - C_{\alpha,a}(t))$  converges weakly to a standard Brownian motion as  $a \rightarrow 0$  in  $D[0, 1]$  under the topology induced by the uniform metric if and only if for each  $\kappa > 0$  we have  $\sigma(\alpha, \kappa\sigma(a) \wedge a) \sim \sigma(\alpha, a)$  as  $a \rightarrow 0$ , where

$$\sigma(\alpha, a)^2 = E[L_{\alpha,a}(1)^2] = \int_{-a}^a y^2 \lambda_L(dy) = \frac{\alpha}{2-\alpha} c_\alpha a^{2-\alpha}, \quad (3.63)$$

$B$  is a standard Brownian motion, and  $D[0, 1]$  denotes the space of real-valued right continuous functions with left limits [27].

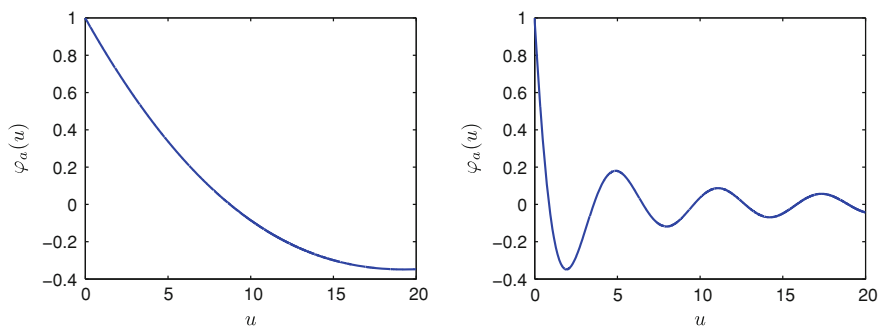
The characteristic function of  $\tilde{L}_{\alpha,a}(t) = \sigma(\alpha, a)B(t) + C_{\alpha,a}(t)$  is

$$\tilde{\varphi}_{\alpha,a}(u; t) = \exp\left[-\frac{(u\sigma(\alpha, a))^2 t}{2} - \lambda_a t(1 - \varphi_a(u))\right], \quad (3.64)$$

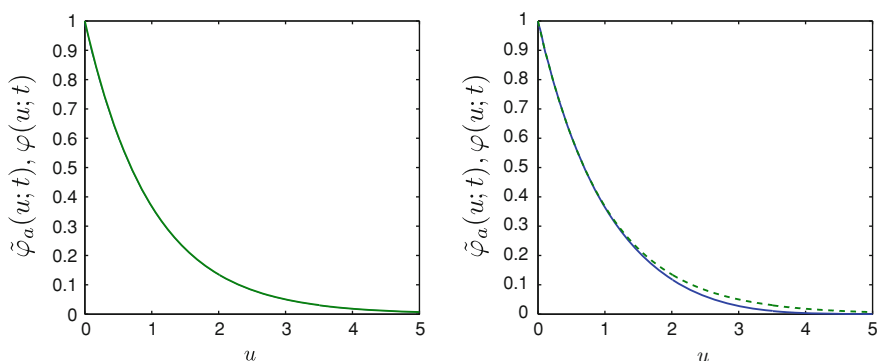
where  $\lambda_a = c_\alpha/a^\alpha$  and  $\varphi_a$  is given by (3.60). Figure 3.8 shows the characteristic functions  $\varphi_a(u)$ ,  $u \geq 0$ , for  $\alpha = 1$ ,  $a = 0.1$  (left panel), and  $a = 1$  (right panel). The difference between these functions is significant. The solid and dotted lines in Fig. 3.9 are the characteristic functions of  $L_\alpha(t)$  and  $\tilde{L}_{\alpha,a}(t)$ , respectively, for  $\alpha = 1$ ,  $t = 1$ ,  $a = 0.1$  (left panel), and  $a = 1$  (right panel). Although the characteristic functions  $\varphi_a$  corresponding to different values of  $a$  differ significantly (Fig. 3.8), the approximate characteristic functions of  $L_\alpha$  depend weakly on  $a$  and are accurate. The characteristic function of  $\tilde{L}_{\alpha,a}(t)$  for  $a = 0.1$  coincides with that of  $L_\alpha(t)$  at the figure scale.

### 3.7.6.4 White Noise Processes

Brownian motion and compound Poisson processes are not m.s. differentiable since their covariance functions at arbitrary times  $s$  and  $t$  are proportional to  $s \wedge t$ . The



**Fig. 3.8** Characteristic functions  $\varphi_a$  in (3.60) with  $\alpha = 1$  for  $a = 0.1$  (left panel) and  $a = 1$  (right panel)



**Fig. 3.9** Characteristic functions of  $L_\alpha(t)$  (solid lines) and  $\tilde{L}_{\alpha,a}(t)$  (dotted lines) for  $\alpha = 1$ ,  $t = 1$ ,  $a = 0.1$  (left panel), and  $a = 1$  (right panel)

covariance function of compound Poisson processes exist if their jumps have finite variance. The samples of the Lévy process are also too rough to be differentiable. Yet, it is common in the applied literature to interpret the derivative of the Brownian motion, compound Poisson, and Lévy processes as white noise processes, and refer to them as Gaussian, Poisson, and Lévy white noise processes, respectively. We denote these processes by

$$\begin{aligned}
 W_G(t) &\stackrel{!}{=} dB(t)/dt && \text{(Gaussian white noise),} \\
 W_P(t) &\stackrel{!}{=} dC(t)/dt && \text{(Poisson white noise), and} \\
 W_L(t) &\stackrel{!}{=} dL_\alpha(t)/dt && \text{(Lévy white noise).}
 \end{aligned} \tag{3.65}$$

Since Gaussian, Poisson, and Lévy white noise processes do not exist, calculations with these processes are formal so that results obtained by these calculations are questionable.



In the following chapters we will see how white noise processes can be incorporated rigorously in the theory of stochastic differential equations and develop practical methods for solving these equations. Relationships between white and physical or colored noise processes will also be examined.

We conclude this brief section with the observation that, while samples of Gaussian and Lévy white noise processes cannot be visualized, samples of Poisson white noise can be drawn. They are sequences of iid random variables  $\{Y_n\}$  arriving at the jump times  $\{T_n\}$  of a Poisson process  $N$ , that is,

$$W_P(t) = \sum_{n=1}^{N(t)} Y_n \delta(t - T_n), \quad t \geq 0, \quad (3.66)$$

where  $\delta(\cdot)$  denotes the Dirac delta function.

## 3.8 Monte Carlo Simulation

Our discussion is limited to Monte Carlo simulation methods for generating samples of stationary Gaussian random functions, translation vector processes, and real-valued non-stationary Gaussian processes based on the spectral representation theorem and an extension of this theorem. These and other topics on Monte Carlo simulation are discussed extensively in [5] (Chap. 5), [28, 29].

### 3.8.1 Stationary Gaussian Random Functions

We (1) construct sequences of random functions depending on finite collections of random variables that approach in some sense target Gaussian/translation random functions and (2) develop algorithms for generating samples of these random functions. Developments are based on the spectral representation in Sect. 3.6.4 showing that weakly stationary random functions can be viewed as superpositions of harmonics with random amplitudes.

#### 3.8.1.1 Stochastic Processes

Let  $X(t)$  be an  $\mathbb{R}^d$ -valued stationary Gaussian process with mean zero, covariance function  $c(\tau) = E[X(t + \tau)X(t)']$ , and spectral density  $s(v) = \{s_{kl}(v)\}$ ,  $k, l = 1, \dots, d$ , with support  $[-\bar{v}, \bar{v}]$ ,  $0 < \bar{v} < \infty$ . Consider a partition  $p_n = (0 = \alpha_0 < \alpha_1 < \dots < \alpha_n = \bar{v})$  of frequency band  $[0, \bar{v}]$ , and set  $\Delta v_r = \alpha_r - \alpha_{r-1}$  and  $v_r = (\alpha_{r-1} + \alpha_r)/2$ ,  $r = 1, \dots, n$ . It is common to use equal frequency intervals in which case  $\Delta v_r = \bar{v}/n$  and  $v_r = (r - 1/2)\bar{v}/n$ ,  $r = 1, \dots, n$ . If the support of  $s(v)$

is not bounded, a cutoff frequency  $0 < \bar{v} < \infty$  needs to be selected such that most of the energy of the process is included in the frequency band  $[0, \bar{v}]$ .

Let  $A_r$  and  $B_r$  be  $\mathbb{R}^d$ -valued Gaussian variables with mean zero and second moments

$$\begin{aligned} E[A_{r,k}A_{p,l}] &= E[B_{r,k}B_{p,l}] = \delta_{rp} \int_{\alpha_{r-1}}^{\alpha_r} g_{k,l}(v) dv \simeq \delta_{rp} g_{k,l}(v_r) \Delta v_r, \\ E[A_{r,k}B_{p,l}] &= -E[B_{r,k}A_{p,l}] = \delta_{rp} \int_{\alpha_{r-1}}^{\alpha_r} h_{k,l}(v) dv \simeq \delta_{rp} h_{k,l}(v_r) \Delta v_r, \end{aligned} \quad (3.67)$$

where  $g(v) = s(v) + s(-v)$  and  $h(v) = -i(s(v) - s(-v))$  (Theorem 3.19). The approximations in (3.67) are satisfactory if the functions  $g_{k,l}(v)$  and  $h_{k,l}(v)$  are nearly constant in the intervals  $(\alpha_{r-1}, \alpha_r)$ .

**Theorem 3.46** *If  $\Delta(p_n) = \max_{1 \leq r \leq n} (\alpha_r - \alpha_{r-1}) \rightarrow 0$  and  $\bar{v} \rightarrow \infty$ , then*

$$X^{(n)}(t) = \sum_{r=1}^n [A_r \cos(v_r t) + B_r \sin(v_r t)] \quad (3.68)$$

*becomes a version of the stationary Gaussian process  $X(t)$ .*

*Proof* Since  $X^{(n)}$  has mean zero and covariance function with entries

$$\begin{aligned} E[X_i^{(n)}(s)X_j^{(n)}(t)] &= \sum_{r=1}^n \left\{ \left[ \int_{\alpha_{r-1}}^{\alpha_r} g_{i,j}(v) dv \right] \cos(v_r(t-s)) \right. \\ &\quad \left. - \left[ \int_{\alpha_{r-1}}^{\alpha_r} h_{i,j}(v) dv \right] \sin(v_r(t-s)) \right\}, \quad i, j = 1, \dots, d, \end{aligned}$$

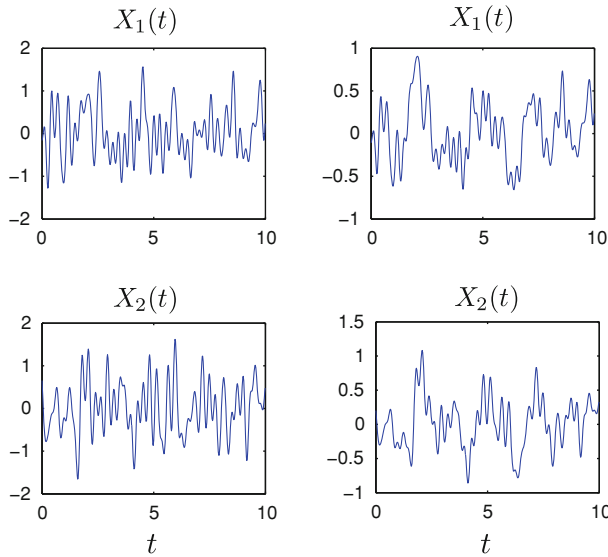
it is weakly stationary for each  $n$ . The process  $X^{(n)}(t)$  is Gaussian as a linear form of the Gaussian variables  $A_r$  and  $B_r$ . If  $d = 1$ , then  $g_{1,1} = g$ ,  $h_{1,1} = 0$ , and  $E[X^{(n)}(s)X^{(n)}(t)] = \sum_{r=1}^n \left[ \int_{\alpha_{r-1}}^{\alpha_r} g(v) dv \right] \cos(v_r(t-s))$ . Since

$$\lim_{n \rightarrow \infty} E[X_i^{(n)}(s)X_j^{(n)}(t)] = \int_0^{\bar{v}} \left[ g_{i,j}(v) \cos(v(t-s)) - h_{i,j}(v) \sin(v(t-s)) \right] dv$$

for a fixed  $\bar{v}$ , the covariance function of  $X^{(n)}$  converges to that of  $X$  by letting  $\bar{v} \rightarrow \infty$  in case the frequency band of  $X$  is not bounded. Since  $X^{(n)}$  and  $X$  are Gaussian processes,  $X^{(n)}$  becomes a version of  $X$  under these limits.  $\blacktriangle$

The statement in Theorem 3.46 justifies the use of samples of  $X^{(n)}$  as a substitute for samples of  $X$ . Additional arguments on the validity of this approximation can be found in [28](Chap. 4) and [29] (Chaps. 1, 2, and 3).

The generation of samples of  $X^{(n)}$  in a time interval  $[0, \tau]$ ,  $\tau > 0$ , involves the following three steps. First, select a cutoff frequency  $0 < \bar{v} < \infty$  in case the frequency band of  $X$  is unbounded, partition the frequency band  $[0, \bar{v}]$ , and specify



**Fig. 3.10** Samples of the coordinates  $X_1$  and  $X_2$  of  $X$  for  $\rho = 0.3$  (left panel) and  $\rho = 0.7$  (right panel)

discrete frequencies  $v_r$ ,  $r = 1, \dots, n$ . Second, generate samples of the Gaussian random variables  $A_r$  and  $B_r$ ,  $r = 1, \dots, n$ , with mean zero and second moments in (3.67). Third, calculate samples of  $X^{(n)}(t)$  in  $[0, \tau]$  from (3.68).

Note that the samples of  $X^{(n)}(t)$  are periodic with period  $2\pi/v_1$  if  $\Delta v_r = \bar{v}/n$  and  $v_r = (r - 1/2)\Delta v_r$  and infinitely differentiable even if the samples of  $X$  are not differentiable. For example, the samples of  $X^{(n)}(t)$  for a stationary Gaussian process  $X(t)$  with exponential correlation are infinitely differentiable although  $X(t)$  is not mean square differentiable.

*Example 3.57* Let  $X(t)$  be an  $\mathbb{R}^2$ -valued stationary Gaussian process with spectral density

$$s_{k,l}(v) = (1 - \rho)\delta_{kl}s_k(v) + \rho s_Z(v) \quad k, l = 1, 2,$$

where  $s_k(v) = 1/(2\bar{v}_k)1(-\bar{v}_k \leq v \leq \bar{v}_k)$ ,  $s_Z(v) = 1/(2\bar{v})1(-\bar{v} \leq v \leq \bar{v})$ , and  $0 < \bar{v}_k, \bar{v} < \infty$ . Figure 3.10 shows samples of the coordinates of  $X$  for  $\bar{v}_k = 25$ ,  $\bar{v} = 5$ ,  $\Delta v_r = \max(\bar{v}_1, \bar{v}_2, \bar{v})/n$ ,  $v_r = (r - 1/2)\Delta v_r$ ,  $r = 1, \dots, n$ ,  $n = 100$ , and two values of  $\rho$ . The frequency content of  $X(t)$  depends strongly on  $\rho$ . The coordinates  $X_1$  and  $X_2$  of  $X$  are nearly in phase for values of  $\rho$  close to unity.  $\diamond$

### 3.8.1.2 Random Fields

Let  $X(t)$ ,  $t \in \mathbb{R}^{d'}$ , be a real-valued, homogeneous Gaussian field with mean zero, covariance function  $c(\tau) = E[X(t + \tau)X(t)]$ , and spectral density  $s(v)$ , where  $t, \tau \in \mathbb{R}^{d'}$  are spatial coordinates and  $v \in \mathbb{R}^{d'}$  is a frequency. Our discussion is limited to real-valued random fields. Extension to vector-valued random fields can be based on arguments similar to those for vector processes.

Let  $D \subset \mathbb{R}^{d'}$  be the support of the spectral density  $s(v)$  of  $X$  assumed to be a bounded set. We (i) take  $D$  to be a bounded rectangle centered at the origin of  $\mathbb{R}^{d'}$ , that is,  $D = \times_{k=1}^{d'} [-\bar{v}_k, \bar{v}_k]$ ,  $\bar{v}_k > 0$ , (ii) partition  $D$  in rectangles  $D_r$ ,  $r = 1, \dots, d'$ , defined by a grid with step  $\Delta v_k = \bar{v}_k/n_k$ ,  $k = 1, \dots, d'$ , where  $n_k \geq 1$  are integers, and (iii) select  $v_r$  at the center of  $D_r$ . Let  $A_r$  and  $B_r$ ,  $r = 1, \dots, n$ , be independent Gaussian variables with mean zero and variance

$$\gamma_r = E[A_r^2] = E[B_r^2] = \int_{D_r} s(v) dv. \quad (3.69)$$

**Theorem 3.47** *The sequence*

$$X^{(n)}(t) = \sum_{r=1}^n [A_r \cos(v_r \cdot t) + B_r \sin(v_r \cdot t)], \quad (3.70)$$

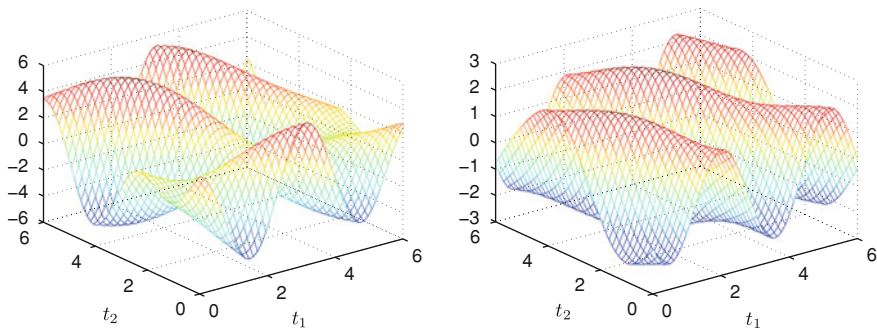
becomes a version of  $X$  in the limit as  $\max_{1 \leq r \leq n} \{\lambda(D_r)\} \rightarrow 0$ ,  $n \rightarrow \infty$ , and  $\min_{1 \leq k \leq d'} \{\bar{v}_k\} \rightarrow \infty$ , where  $v_r \cdot t = \sum_{k=1}^{d'} v_{r,k} t_k$ ,  $v_r = (v_{r,1}, \dots, v_{r,d'})$ ,  $\lambda(D_r)$  denotes the Lebesgue measure of  $D_r$ , and  $t = (t_1, \dots, t_{d'})$ .

*Proof* The sequence of random fields  $X^{(n)}$  are Gaussian with mean zero and covariance functions

$$\begin{aligned} E[X^{(n)}(s)X^{(n)}(t)] &= \sum_{r,p=1}^n E[(A_r \cos(v_r \cdot s) + B_r \sin(v_r \cdot s))(A_p \cos(v_p \cdot t) + B_p \sin(v_p \cdot t))] \\ &= \sum_{r=1}^n \gamma_r (\cos(v_r \cdot s) \cos(v_r \cdot t) + \sin(v_r \cdot s) \sin(v_r \cdot t)) = \sum_{r=1}^n \gamma_r \cos(v_r \cdot (t - s)), \end{aligned}$$

so that they are homogeneous for each  $n \geq 1$ . Considerations as in the proof of Theorem 3.46 show that  $E[X^{(n)}(t)X^{(n)}(s)]$  converges to the covariance function of  $X$  as  $n \rightarrow \infty$  and the boundaries of  $D$  are extended to  $\mathbb{R}^{d'}$ , in case  $D$  is not bounded. Hence,  $X^{(n)}$  becomes a version of  $X$  as the partition of  $D$  is refined and the frequency band is increased indefinitely.  $\blacktriangle$

The generation of samples of  $X^{(n)}$  in a subset  $S$  of  $\mathbb{R}^{d'}$  involves the following three steps. First, specify cutoff frequencies  $\bar{v}_k > 0$ ,  $k = 1, \dots, d'$ , in case the frequency band of  $X$  is unbounded, partition  $D = \times_{k=1}^{d'} [-\bar{v}_k, \bar{v}_k]$  in  $n$  rectangles, and select



**Fig.3.11** Samples of  $X$  for two discrete spectral densities

interior points  $v_r$ ,  $r = 1, \dots, n$ , in these rectangles. Second, generate samples of the Gaussian random variables  $A_r$  and  $B_r$ ,  $r = 1, \dots, n$ , with mean zero and second moments in (3.69). Third, calculate samples of  $X^{(n)}(t)$ ,  $t \in S$ , from (3.70).

*Example 3.58* Let  $X(t)$ ,  $t \in \mathbb{R}^2$ , be a real-valued homogeneous Gaussian field defined by

$$X(t) = \sum_{r=1}^n [A_r \cos(v_r \cdot t) + B_r \sin(v_r \cdot t)],$$

where  $A_r$  and  $B_r$  are independent Gaussian variables with mean zero and variance  $E[A_r^2] = E[B_r^2] = \gamma_r$ . Since the field  $X(t)$  has the form in (3.70), it can be used directly to produce samples. Figure 3.11 shows two samples of  $X$  for  $n = 6$ ,  $v_1 = (1, 2)$ ,  $v_2 = (2, 1)$ ,  $v_3 = (2, 2)$ ,  $v_4 = -v_1$ ,  $v_5 = -v_2$ , and  $v_6 = -v_3$ . Recall that spectral densities  $s(v)$  of real-valued homogeneous random fields must satisfy the condition  $s(v) = s(-v)$  (Theorem 3.8). The left sample is for  $\gamma_r = 1$ ,  $r = 1, \dots, 6$ . The right sample corresponds to  $\gamma_1 = \gamma_4 = 1$  and  $\gamma_r = 0.01$  for  $r \neq 1, 4$ , and has a dominant wave with frequency  $v_1 = (1, 2)$ .  $\diamond$

### 3.8.2 Translation Vector Processes

Our objectives are to (1) construct translation models  $X^{(T)}(t)$  for  $\mathbb{R}^d$ -valued stationary non-Gaussian processes  $X(t)$  that match both the marginal distributions and the covariance functions of these processes and (2) develop a Monte Carlo algorithm for generating samples of  $X^{(T)}(t)$ . Translation models were introduced in Sect. 3.7.2.

The first objective cannot always be achieved since translation models with the required properties may not exist ([4], Sect. 3.1.1), in which case we have three options [16]: match the marginal distributions and approximate the covariance function, approximate the marginal distributions and match the covariance function, and

approximate both the marginal distributions and the covariance functions. We discuss the first option.

Let  $G(t)$  be an  $\mathbb{R}^d$ -valued weakly stationary process whose coordinates  $G_k(t)$ ,  $k = 1, \dots, d$ , have mean 0, variance 1, covariance functions  $\rho_{kl}(\tau) = E[G_k(t + \tau)G_l(t)]$ , and spectral densities  $s_{kl}(v)$ ,  $k, l = 1, \dots, n$ . The covariance and spectral density functions are related by (Theorem 3.7)

$$\rho_{k,l}(\tau) = \int_{-\infty}^{\infty} e^{iv\tau} s_{k,l}(v) dv \quad \text{and} \quad s_{k,l}(v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iv\tau} \rho_{k,l}(\tau) d\tau. \quad (3.71)$$

**Theorem 3.48** *The spectral densities  $\{s_{k,l}(v)\}$  in (3.71) are such that (1)  $s_{k,k}(v)$ ,  $v \in \mathbb{R}$ , are real-valued, even, positive functions, (2) the matrix  $\{s_{k,l}(v)\}$  of spectral densities is Hermitian, that is,  $s_{k,l}(v) = s_{l,k}(v)^*$  for all  $v \in \mathbb{R}$ , (3)  $s_{k,l}(v) = s_{k,l}(-v)^*$ , (4)  $|s_{k,l}(v)|^2 \leq s_{k,k}(v)s_{l,l}(v)$ , and (5)  $s_{k,k}(v) + s_{l,l}(v) + 2\Re[s_{k,l}(v)] \geq 0$ ,  $k \neq l$ .*

*Proof* The first three properties and (3.71) follow from Theorem 3.17. For example,  $s_{k,l}(v) = \int_{-\infty}^{\infty} e^{-iv\tau} \rho_{k,l}(\tau) d\tau / (2\pi) = \int_{-\infty}^{\infty} e^{iv\tau} \rho_{k,l}(-\tau) (-d\tau) / (2\pi)$  by the change of variable  $\tau \mapsto -\tau$  so that  $s_{k,l}(v) = \int_{-\infty}^{\infty} e^{iv\tau} \rho_{l,k}(\tau) d\tau / (2\pi) = s_{l,k}(v)^*$  since  $\rho_{k,l}(\tau) = \rho_{l,k}(-\tau)$ .

The proof of the fourth property can be found in [8] (Sect. 8.1) or [5] (Sect. 3.7.2.4). We prove the fifth property. The real-valued and complex-valued processes  $Y_1(t) = G_k(t) + G_l(t)$  and  $Y_2(t) = iG_k(t) + G_l(t)$  are weakly stationary with mean 0 and covariance functions

$$\begin{aligned} c_1(\tau) &= E[Y_1(t + \tau)Y_1(t)^*] = \rho_{k,k}(\tau) + \rho_{l,l}(\tau) + \rho_{k,l}(\tau) + \rho_{l,k}(\tau) \quad \text{and} \\ c_2(\tau) &= E[Y_2(t + \tau)Y_2(t)^*] = \rho_{k,k}(\tau) + \rho_{l,l}(\tau) + i\rho_{k,l}(\tau) - i\rho_{l,k}(\tau), \end{aligned}$$

respectively. Let  $s_p(v) = \int_{-\infty}^{\infty} e^{-iv\tau} c_p(\tau) d\tau / (2\pi)$  denote the spectral density functions of  $Y_p(t)$ ,  $p = 1, 2$ . We have

$$\begin{aligned} s_1(v) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iv\tau} [\rho_{k,k}(\tau) + \rho_{l,l}(\tau) + \rho_{k,l}(\tau) + \rho_{l,k}(\tau)] d\tau \\ &= s_{k,k}(v) + s_{l,l}(v) + s_{k,l}(v) + s_{l,k}(v) \end{aligned}$$

by (3.71), so that  $s_1(v) = s_{k,k}(v) + s_{l,l}(v) + 2\Re[s_{k,l}(v)]$  by property (3). Since  $Y_1(t)$  is a real-valued process, its spectral density must be positive, that is,  $s_{k,k}(v) + s_{l,l}(v) + 2\Re[s_{k,l}(v)] \geq 0$  for all  $v \in \mathbb{R}$ . This inequality provides constraints on the possible values of spectral densities  $s_{k,l}(v)$  relative to those of  $s_{k,k}(v)$  and  $s_{l,l}(v)$ . The spectral density of the complex-valued stochastic process  $Y_2(t)$  is  $s_2(v) = s_{k,k}(v) + s_{l,l}(v) + i(s_{k,l}(v) - s_{l,k}(v)) = s_{k,k}(v) + s_{l,l}(v) - 2\Im[s_{k,l}(v)]$ ,  $v \in \mathbb{R}$ , and provides no additional constraints for the spectral densities of  $G(t)$ .  $\blacktriangle$

Let  $X(t)$  be an  $\mathbb{R}^d$ -valued non-Gaussian stationary process whose coordinates  $X_k(t)$  have means  $E[X_k(t)] = 0$ , variances  $\text{Var}[X_k(t)] = E[X_k(t)^2] = 1$ , distributions  $F_k(x) = P(X_k(t) \leq x)$ , and covariance functions  $\xi_{k,l}(\tau) = E[X_k(t + \tau)X_l(t)]$ ,  $k, l = 1, \dots, d$ . The assumption that the coordinates of  $X(t)$  have zero

means and unit variances is not restrictive since any stationary process can be modified to have these properties by shifting and scaling.

Let  $X^{(T)}(t)$  be an  $\mathbb{R}^d$ -valued translation model for  $X(t)$  defined by

$$X_k^{(T)}(t) = F_k^{-1} \circ \Phi(G_k(t)) = h_k(G_k(t)), \quad k = 1, \dots, d, \quad (3.72)$$

where  $\Phi$  denotes the distribution of  $N(0, 1)$ ,  $h_k = F_k^{-1} \circ \Phi$ , and  $\{G_k(t)\}$  are the coordinates of an  $\mathbb{R}^d$ -valued stationary Gaussian process  $G(t)$  with  $E[G_k(t)] = 0$ ,  $E[G_k(t)^2] = 1$ ,  $\rho_{k,l}(\tau) = E[G_k(t + \tau)G_l(t)]$ , and  $s_{k,l}(v) = \int_{-\infty}^{\infty} e^{-iv\tau} \rho_{k,l}(\tau) d\tau / (2\pi)$ ,  $k, l = 1, \dots, d$ .

The marginal distributions of  $X^{(T)}(t)$  coincide with those of  $X(t)$  irrespective of the spectral densities  $\{s_{k,l}(v)\}$  or equivalently, the covariance functions  $\rho_{k,l}(\tau)$ , of  $G(t)$  since  $P(X_{T,k}(t) \leq x) = P(G_k(t) \leq \Phi^{-1} \circ F_k(x)) = F_k(x)$ ,  $k = 1, \dots, d$ . On the other hand, the covariance functions,

$$\begin{aligned} \xi_{k,l}^{(T)}(\tau) &= E[X_k^{(T)}(t + \tau)X_l^{(T)}(t)] = E[h_k(G_k(t + \tau))h_l(G_l(t))] \\ &= \int_{\mathbb{R}^2} h_k(u)h_l(v)\phi(u, v; \rho_{kl}(\tau)) du dv, \end{aligned} \quad (3.73)$$

of  $X^{(T)}(t)$  depend on both the second moment properties of  $G(t)$  and the mapping in (3.72). The function  $\phi(u, v; \rho)$  in (3.73) denotes the joint density of a standard bivariate Gaussian vector with correlation coefficient  $\rho$ .

Consider the case in which there is no translation model matching the marginal distributions and covariance functions of a target non-Gaussian vector process  $X(t)$ . We construct a translation process  $X^{(T)}(t)$  that has the same marginal distributions as  $X(t)$ . The coordinates  $\{G_k(t)\}$  of the Gaussian image  $G(t)$  of  $X^{(T)}(t)$  have mean 0, variance 1, and their covariance functions are such that the covariance functions of  $X^{(T)}(t)$  are as close as possible to those of  $X(t)$  in some sense.

Suppose the spectral densities  $\{s_{k,l}(v)\}$  of  $G(t)$  have non-zero ordinates in a frequency band  $(-\bar{v}, \bar{v})$ ,  $0 < \bar{v} < \infty$ . Let  $n > 1$  be an integer,  $\Delta v = \bar{v}/n$ ,  $v_r = (r - 1/2)\Delta v$  and  $v_{-r} = -v_r$  for  $r = 1, \dots, n$ , and approximate  $\{s_{k,l}(v)\}$  by the discrete spectral densities

$$\tilde{s}_{k,l}(v) = \sum_{r=-n, r \neq 0}^n s_{k,l}^{(r)} \delta(v - v_r), \quad (3.74)$$

where  $\{s_{k,l}^{(r)}\}$  and  $\{s_{k,l}^{(-r)}\}$  are unknown spectral ordinates associated with frequencies  $\{v_r\}$  and  $\{v_{-r}\}$ . Note that  $\{\tilde{s}_{k,l}(v)\}$  must have the properties stated in Theorem 3.48. If the spectral densities  $\{\tilde{s}_{k,l}(v)\}$  satisfy the conditions of this theorem, then

$$\begin{aligned} \tilde{\rho}_{k,l}(\tau) &= \int_{-\infty}^{\infty} e^{iv\tau} \tilde{s}_{k,l}(v) dv = \int_{-\infty}^{\infty} e^{iv\tau} \sum_{r=-n, r \neq 0}^n s_{k,l}^{(r)} \delta(v - v_r) dv \\ &= \sum_{r=-n, r \neq 0}^n e^{iv_r\tau} s_{k,l}^{(r)} = 2 \sum_{r=1}^n \left[ \Re[s_{k,l}^{(r)}] \cos(v_r\tau) - \Im[s_{k,l}^{(r)}] \sin(v_r\tau) \right] \end{aligned} \quad (3.75)$$

are covariance functions. Since  $\{\tilde{\rho}_{k,l}(\tau)\}$  depend only on the ordinates  $\{s_{k,l}^{(r)}\}$  of the discrete spectra of  $G(t)$ , so are the covariances  $\{\tilde{\xi}_{k,l}^{(T)}(\tau) = E[\tilde{X}_k^{(T)}(t + \tau)\tilde{X}_l^{(T)}(t)]\}$  of the corresponding translation model  $\tilde{X}^{(T)}(t)$  of  $X(t)$  defined by

$$\tilde{X}_k^{(T)}(t) = F_k^{-1} \circ \Phi(\tilde{G}_k(t)) = h_k(\tilde{G}_k(t)), \quad k = 1, \dots, d, \quad (3.76)$$

where  $\tilde{G}(t) = (\tilde{G}_1(t), \dots, \tilde{G}_d(t))$  is an  $\mathbb{R}^d$ -valued stationary Gaussian process with covariance functions  $\{\tilde{\rho}_{k,l}(\tau)\}$  and spectra  $\{\tilde{s}_{k,l}(\nu)\}$ . Let

$$s_{\text{vect}} = \{s_{k,k}^{(r)}, s_{l,l}^{(r)}, \Re[s_{k,l}^{(r)}], \Im[s_{k,l}^{(r)}], k, l = 1, \dots, d, r = 1, \dots, n\} \quad (3.77)$$

be a vector containing the unknown parameters in the definition of  $\tilde{G}(t)$ .

**Theorem 3.49** *The optimal Gaussian process  $\tilde{G}(t)$  is given by  $s_{\text{vect}}$  in (3.77) that minimizes the objective function*

$$e(s_{\text{vect}}) = \max_{\tau \geq 0} \left[ \sum_{k,l=1}^d \left( \tilde{\xi}_{k,l}^{(T)}(\tau) - \xi_{k,l}(\tau) \right)^2 \right] \quad (3.78)$$

under the constraints

$$\begin{aligned} s_{k,k}^{(r)} &\geq 0, \quad \sum_{r=1}^n s_{k,k}^{(r)} = 1/2, \quad |s_{k,l}^{(r)}|^2 \leq s_{k,k}^{(r)} s_{l,l}^{(r)}, \quad k \neq l, \quad \text{and} \\ s_{k,k}^{(r)} + s_{l,l}^{(r)} + 2\Re[s_{k,l}^{(r)}] &\geq 0, \quad k \neq l, \end{aligned} \quad (3.79)$$

for all  $k, l = 1, \dots, d$  and  $r = 1, \dots, n$ .

The calculation of the covariance functions  $\tilde{\xi}_{k,l}^{(T)}(\tau)$  of the translation model  $\tilde{X}^{(T)}(t)$  with Gaussian image  $\tilde{G}(t)$  involves numerical evaluations of double integrals, which may render the optimization algorithm in (3.78) and (3.79) unfeasible. This difficulty can be resolved by precalculating and storing the mappings  $\rho \mapsto \xi$  between correlation coefficients of standard bivariate Gaussian vectors  $(G_k, G_l)$  and those of bivariate non-Gaussian vectors  $(F_k^{-1} \circ \Phi(G_k), F_l^{-1} \circ \Phi(G_l))$ . These mappings are monotonically increasing ([4], Sect. 3.1.1) and can be used to find covariances  $\tilde{\xi}_{k,l}^{(T)}(\tau)$  from  $\tilde{\rho}_{k,l}(\tau)$  and evaluate the objective function in (3.78).

Once a translation model  $\tilde{X}^{(T)}(t)$  has been constructed, the following algorithm can be used for sample generation. First, specify a cutoff frequency  $\bar{\nu} > 0$  for the coordinates of  $G(t)$  and select discrete frequencies in the frequency band considered for  $G(t)$ . Second, generate samples of  $\tilde{G}(t)$  defined by

$$\tilde{G}(t) = \sum_{r=-n, r \neq 0}^n C_r e^{i\nu_r t}, \quad (3.80)$$



where the coefficients  $\{C_r\}$  are  $\mathbb{C}^d$ -valued Gaussian variables such that  $E[\tilde{G}(t)] = 0$  and  $E[\tilde{G}_k(s)\tilde{G}(t)] = \tilde{\rho}_{kl}(t-s)$  (Example 3.15). The generation of samples of  $\tilde{G}(t)$  can be based on the algorithm outlined in Sect. 3.8.1.2 using the representation in (3.68), which is an alternative form of (3.80).

*Example 3.59* Let  $G(t) = (G_1(t), G_2(t))$  be a bivariate stationary Gaussian process defined by  $G_k(t) = \sqrt{1-\theta}Z_k(t) + \sqrt{\theta}Z(t)$ ,  $k = 1, 2$ , where  $\theta \in [0, 1]$  and  $Z_k(t)$ ,  $k = 1, 2$ , and  $Z(t)$  are independent stationary Gaussian processes with zero means and spectral densities  $s_{Z_k}(v) = 1(-\bar{v}_{Z_k} < v < \bar{v}_{Z_k})/(2\bar{v}_{Z_k})$ ,  $0 < \bar{v}_{Z_k} < \infty$ , and  $s_Z(v) = 1(-\bar{v}_Z < v < \bar{v}_Z)/(2\bar{v}_Z)$ ,  $0 < \bar{v}_Z < \infty$ . The spectral densities and covariance functions of  $G(t)$  are

$$\begin{aligned} s_{k,l}(v) &= (1-\theta)\delta_{kl}s_{Z_k}(v) + \theta s_Z(v) \quad \text{and} \\ \rho_{k,l}(\tau) &= (1-\theta)\delta_{kl}\frac{\sin(\bar{v}_{Z_k}\tau)}{\bar{v}_{Z_k}\tau} + \theta\frac{\sin(\bar{v}_Z\tau)}{\bar{v}_Z\tau} \end{aligned} \quad (3.81)$$

Define an  $\mathbb{R}^2$ -valued translation non-Gaussian process  $X^{(T)}(t)$  by (3.72) with

$$\begin{aligned} F_1(x) &= \Phi(\ln(x-a)/\sigma), \quad a, \sigma > 0, \quad x > a, \\ F_2(x) &= \Phi(\text{sign}(x)|x|^{1/3}), \quad x \in \mathbb{R}, \end{aligned} \quad (3.82)$$

and suppose that  $X(t) = X^{(T)}(t)$  is the target non-Gaussian process.

The covariance functions of the coordinates of  $X(t)$  are related by ([4], Sect. 3.1.1)

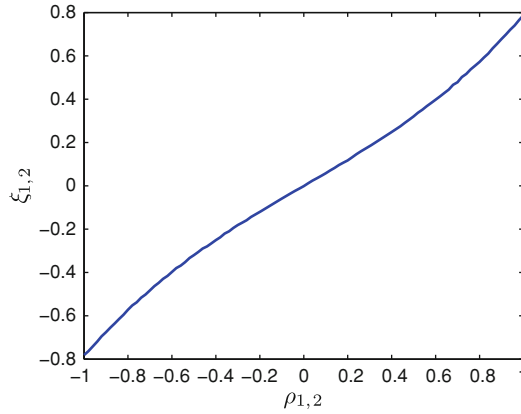
$$\begin{aligned} \xi_{1,1}(\tau) &= (1 - e^{\sigma^2\rho_{1,1}(\tau)})/(1 - e^{\sigma^2}) \quad \text{and} \\ \xi_{2,2}(\tau) &= \frac{1}{5}\rho_{2,2}(\tau)(3 + 2\rho_{2,2}(\tau)^2). \end{aligned} \quad (3.83)$$

The relationship between  $\rho_{1,2}(\tau)$  and  $\xi_{1,2}(\tau)$  cannot be obtained analytically. Figure 3.12 shows an estimate of this relationship calculated by Monte Carlo simulation from (3.73) using  $10^6$  samples of bivariate standard Gaussian vectors with correlation coefficients  $\rho_{1,2} \in [-1, 1]$ .

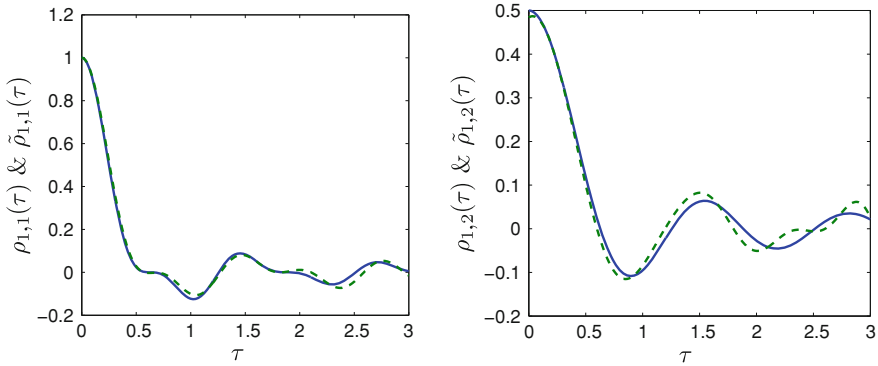
Consider a discrete representation  $\tilde{s}_{k,l}(v)$  of the spectral densities  $s_{k,l}(v)$  of  $G(t)$  as given by (3.74). The representation depends on the spectral ordinates  $s_{k,k}^{(r)} \in \mathbb{R}$ ,  $k = 1, 2$ ,  $\Re[s_{1,2}^{(r)}] \in \mathbb{R}$ , and  $\Im[s_{1,2}^{(r)}] \in \mathbb{R}$ ,  $r = 1, \dots, n$ . Let, as in (3.77),

$$s_{\text{vect}} = \{s_{1,1}^{(1)}, \dots, s_{1,1}^{(n)}, s_{2,2}^{(1)}, \dots, s_{2,2}^{(n)}, \Re[s_{1,2}^{(1)}], \dots, \Re[s_{1,2}^{(n)}], \Im[s_{1,2}^{(1)}], \dots, \Im[s_{1,2}^{(n)}]\} \quad (3.84)$$

be a vector in  $\mathbb{R}^{4n}$  including the unknown parameters in the discrete version of the spectral densities of  $G(t)$ . Our objective is to find an optimal  $s_{\text{vect}}$  in the sense that it minimizes the objective function in (3.78). The covariance functions  $\tilde{\xi}_{k,l}^{(T)}(\tau)$  in (3.78) are images of the covariance functions  $\tilde{\rho}_{k,l}(\tau)$  given by (3.75), and can be obtained from (3.78) and Fig. 3.12.



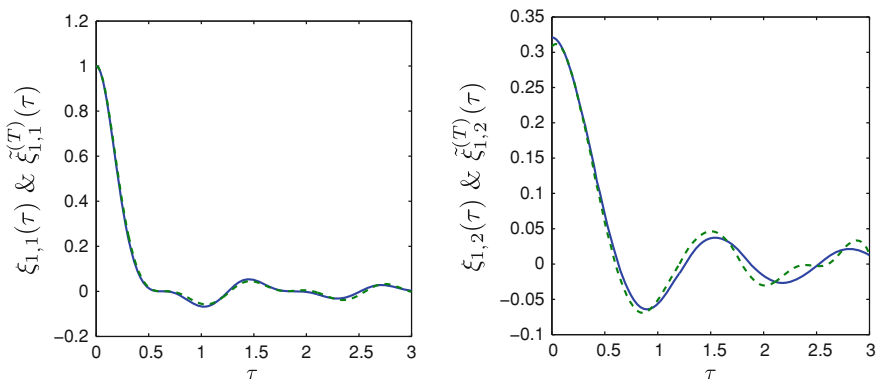
**Fig. 3.12** An estimate of the relationship between correlation coefficients  $\rho_{1,2}$  of  $(G_1, G_2)$  and  $\xi_{1,2}$  of  $(F_1^{-1}\Phi(G_1), F_2^{-1}\Phi(G_2))$



**Fig. 3.13** Target and optimal covariance functions in the Gaussian space,  $\rho_{k,l}(\tau)$  and  $\tilde{\rho}_{k,l}(\tau)$

Numerical results have been obtained for  $\bar{v}_{Z_k} = 10$ ,  $k = 1, 2$ ,  $\bar{v}_Z = 5$ ,  $n=15$ ,  $a = 1$ ,  $\sigma = 1$ , and  $\theta = 0.5$ . Figure 3.13 shows with solid and dotted lines the target covariance functions  $\rho_{k,l}(\tau)$  and the optimal covariance functions  $\tilde{\rho}_{k,l}(\tau)$  in the Gaussian space. Target and translation covariance functions,  $\xi_{k,l}(\tau)$  and  $\tilde{\xi}_{k,l}^{(T)}(\tau)$ , are shown with solid and dotted lines in Fig. 3.14. The optimal covariance functions nearly coincide with the target covariance functions, an expected result since  $X(t)$  is a translation process so that it can be mapped exactly into a Gaussian process.  $\diamond$

The discrete approximation of the spectral densities  $\{s_{k,l}(v)\}$  of  $G(t)$  in (3.74) is feasible for stochastic processes. However, it is impractical for random fields defined on  $\mathbb{R}^d$ ,  $d \geq 2$ , since the dimension of the vector  $s_{\text{vect}}$  of unknown parameters in (3.77) is excessive. For translation random fields, it is convenient to approximate the spectral densities of their Gaussian images by finite sums of specified functions weighted by unknown coefficients. The resulting representations of the spectral densities must satisfy conditions as those in Theorem 3.48.



**Fig.3.14** Target and optimal covariance functions in the translation space,  $\xi_{k,l}^{(T)}(\tau)$  and  $\tilde{\xi}_{k,l}(\tau)$

### 3.8.3 Non-Stationary Gaussian Processes

Let  $X$  be a real-valued non-stationary Gaussian process with mean 0, covariance function  $c(s, t) = E[X(s)X(t)]$ , and generalized spectral density  $s(v, \eta)$ , where ([30], Sect. 12-2)

$$\begin{aligned} s(v, \eta) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} c(s, t) e^{-i(vs - \eta t)} ds dt \\ c(s, t) &= \int_{\mathbb{R}^2} s(v, \eta) e^{i(vs - \eta t)} dv d\eta. \end{aligned} \quad (3.85)$$

We construct approximations for the covariance function  $c(s, t)$  by (1) truncating the frequency band of this process, that is, replacing  $s(v, \eta)$  with  $s(v, \eta)1((v, \eta) \in I(\bar{v}, \bar{\eta}))$ , where  $I(\bar{v}, \bar{\eta}) = [-\bar{v}, \bar{v}] \times [-\bar{\eta}, \bar{\eta}]$ ,  $0 < \bar{v}, \bar{\eta} < \infty$ , is a bounded rectangle such that  $\int_{I(\bar{v}, \bar{\eta})} s(v, \eta) e^{i(vs - \eta t)} dv d\eta \simeq \int_{\mathbb{R}^2} s(v, \eta) e^{i(vs - \eta t)} dv d\eta = c(s, t)$  at all times and (2) partitioning  $I(\bar{v}, \bar{\eta})$  in sufficiently small subsets such that  $s(v, \eta)$  is nearly constant in each of these subsets.

Resulting approximations of  $c(s, t)$  are used to construct a sequence of processes converging to  $X$ . Denote by  $\tilde{c}(s, t)$  an approximation of  $c(s, t)$  corresponding to the truncated spectral density  $s(v, \eta)1((v, \eta) \in I(\bar{v}, \bar{\eta}))$ , that is,

$$\tilde{c}(s, t) = \int_{I(\bar{v}, \bar{\eta})} s(v, \eta) e^{i(vs - \eta t)} dv d\eta, \quad (s, t) \in \mathbb{R}^2. \quad (3.86)$$

Let  $J(\bar{\tau}, \bar{\rho}) = [-\bar{\tau}, \bar{\tau}] \times [-\bar{\rho}, \bar{\rho}] \in \mathbb{R}^2$ ,  $0 < \bar{\tau}, \bar{\rho} < \infty$ , be a bounded rectangle in the space of temporal coordinates, and approximate  $s(v, \eta)$  by

$$\tilde{s}(v, \eta) = \frac{1}{(2\pi)^2} \int_{J(\bar{\tau}, \bar{\rho})} c(s, t) e^{-i(vs - \eta t)} ds dt, \quad (v, \eta) \in \mathbb{R}^2. \quad (3.87)$$

**Theorem 3.50** *If  $c(s, t)$  is continuous, has bounded partial derivatives  $\partial c(s, t)/\partial s$  and  $\partial c(s, t)/\partial t$ , and the mixed partial derivative  $\partial^2 c(s, t)/\partial s \partial t$  exists in  $J(\bar{\tau}, \bar{\rho})$ , then the Fourier series of  $c(s, t)$  restricted to the bounded rectangle  $J(\bar{\tau}, \bar{\rho})$  converges to  $c(s, t)$  at all interior points of this rectangle.*

*Proof* The proof can be found in [31], Sect. 7.3. The meaning of this theorem is that the partial sums,

$$\begin{aligned}\sigma_{p,q}(s, t) &= \sum_{k=-p}^p \sum_{l=-q}^q \zeta_{k,l} e^{i\pi(ks/\bar{\tau}-lt/\bar{\rho})}, \quad \text{where} \\ \zeta_{k,l} &= \frac{1}{4\bar{\tau}\bar{\rho}} \int_{J(\bar{\tau}, \bar{\rho})} c(s, t) e^{-i\pi(ks/\bar{\tau}-lt/\bar{\rho})} ds dt,\end{aligned}\quad (3.88)$$

of the double Fourier series of  $c(s, t)$  restricted to  $J(\bar{\tau}, \bar{\rho})$  converge to  $c(s, t)$  as  $p, q \rightarrow \infty$  at all interior points of  $J(\bar{\tau}, \bar{\rho})$ . Hence, the error  $|\sigma_{p,q}(s, t) - c(s, t)|$  of the approximation  $\sigma_{p,q}(s, t)$  of  $c(s, t)$  can be made as small as desired in  $J(\bar{\tau}, \bar{\rho})$  by increasing  $p$  and  $q$ .  $\blacktriangle$

**Theorem 3.51** *If  $s(v, \eta)$  is absolutely integrable in  $\mathbb{R}^2$ , then*

$$|\tilde{c}(s, t) - c(s, t)| = \left| \int_{I(\bar{v}, \bar{\eta})^c} s(v, \eta) e^{i(vs-\eta t)} dv d\eta \right| \leq \int_{I(\bar{v}, \bar{\eta})^c} |s(v, \eta)| dv d\eta, \quad (3.89)$$

*so that the approximate covariance function  $\tilde{c}(s, t)$  of  $X(t)$  converges to  $c(s, t)$  as  $\bar{v}, \bar{\eta} \rightarrow \infty$ . If  $c(s, t)$  is absolutely integrable in  $\mathbb{R}^2$ , then*

$$\begin{aligned}|s(v, \eta) - \tilde{s}(v, \eta)| &= \frac{1}{(2\pi)^2} \left| \int_{J(\bar{\tau}, \bar{\rho})^c} c(s, t) e^{-i(vs-\eta t)} ds dt \right| \\ &\leq \frac{1}{(2\pi)^2} \int_{J(\bar{\tau}, \bar{\rho})^c} |c(s, t)| ds dt\end{aligned}\quad (3.90)$$

*so that  $\tilde{s}(v, \eta)$  converges to  $s(v, \eta)$  as  $\bar{\tau}, \bar{\rho} \rightarrow \infty$ .*

*Proof* The bound on the discrepancy between  $\tilde{c}(s, t)$  and  $c(s, t)$  in (3.89) implies  $\lim_{\bar{v}, \bar{\eta} \rightarrow \infty} \tilde{c}(s, t) = c(s, t)$  at each  $(s, t)$  since  $|s(v, \eta)|$  is integrable in  $\mathbb{R}^2$  by assumption. The requirement that  $s(v, \eta)$  is absolutely integrable in  $\mathbb{R}^2$  is stronger than  $c(t, t) < \infty$  at all times since  $c(t, t) = \left| \int_{\mathbb{R}^2} s(v, \eta) e^{i(v-\eta)t} dv d\eta \right| \leq \int_{\mathbb{R}^2} |s(v, \eta)| dv d\eta$ . For the special case in which  $s(v, \eta)$  has a bounded support, the function  $\tilde{c}(s, t)$  coincides with the target covariance function  $c(s, t)$  provided  $I(\bar{v}, \bar{\eta})$  includes the support of  $s(v, \eta)$ .

The bound on  $|s(v, \eta) - \tilde{s}(v, \eta)|$  in (3.90) implies the stated convergence under the assumption that  $c(s, t)$  is absolutely integrable in  $\mathbb{R}^2$ .  $\blacktriangle$

Let  $I(\bar{v}, \bar{\eta})$  be a bounded rectangle as in (3.86) that may or may not include the support of  $s(v, \eta)$ . For arbitrary integers  $m, n \geq 1$ , set  $\Delta v = \bar{v}/m$ ,  $\Delta \eta = \bar{\eta}/n$ ,  $v_k = k\Delta v$  for  $k = 0, 1, \dots, m$ ,  $v_{-k} = -v_k$  for  $k = 1, 2, \dots, m$ ,  $\eta_l = l\Delta \eta$  for  $l = 0, 1, \dots, n$ ,  $\eta_{-l} = -\eta_l$  for  $l = 1, \dots, n$ , and

$$\begin{aligned}
I_{0,0} &= [-\Delta v/2, \Delta v/2) \times [-\Delta \eta/2, \Delta \eta/2), \\
I_{k,0} &= [v_k - \Delta v/2, v_k + \Delta v/2) \times [-\Delta \eta/2, \Delta \eta/2), \quad k \in \{-m, \dots, -1, 1, \dots, m\} \\
I_{0,l} &= [-\Delta v/2, \Delta v/2) \times [\eta_l - \Delta \eta/2, \eta_l + \Delta \eta/2), \quad l \in \{-n, \dots, -1, 1, \dots, n\} \\
I_{k,l} &= [v_k - \Delta v/2, v_k + \Delta v/2) \times [\eta_l - \Delta \eta/2, \eta_l + \Delta \eta/2), \\
(k, l) &\in \{-m, \dots, -1, 1, \dots, m\} \times \{-n, \dots, -1, 1, \dots, n\}.
\end{aligned} \tag{3.91}$$

The rectangles  $\{I_{k,l}\}$  partition  $I(\bar{v}, \bar{\eta})$ , that is, they are disjoint sets such that  $I(\bar{v}, \bar{\eta}) = \bigcup_{k=-m}^m \bigcup_{l=-n}^n I_{k,l}$ , so that  $\tilde{c}(s, t)$  in (3.86) becomes

$$\tilde{c}(s, t) = \sum_{k=-m}^m \sum_{l=-n}^n \int_{I_{k,l}} s(v, \eta) e^{i(vs - \eta t)} dv d\eta, \tag{3.92}$$

and can be approximated by

$$\tilde{c}_{m,n}(s, t) = \sum_{k=-m}^m \sum_{l=-n}^n s_{k,l} e^{i(v_k s - \eta_l t)}, \tag{3.93}$$

where  $s_{k,l} = \int_{I_{k,l}} s(v, \eta) dv d\eta \simeq s(v_k, \eta_l) \Delta v \Delta \eta$ . The latter approximation holds if  $s(v, \eta)$  is nearly constant in  $I_{k,l}$ . Note that  $\tilde{c}_{m,n}(s, t)$  in (3.93) corresponds to the discrete approximation  $\sum_{k=-m}^m \sum_{l=-n}^n s_{k,l} \delta(v - v_k) \delta(\eta - \eta_l)$  of the truncated generalized spectral density  $s(v, \eta) 1((v, \eta) \in I(\bar{v}, \bar{\eta}))$ .

**Theorem 3.52** *If  $c(s, t)$  and  $s(v, \eta)$  satisfy the conditions in Theorems 3.50 and 3.51, and  $s(v, \eta)$  is continuous, the discrepancy  $|\tilde{c}_{m,n}(s, t) - \sigma_{p,q}(s, t)|$  between  $\tilde{c}_{m,n}(s, t)$  and  $\sigma_{p,q}(s, t)$  can be made arbitrarily small in  $J(\bar{\tau}, \bar{\rho})$  for sufficiently large truncation levels in both frequency and time domains as  $m, n \rightarrow \infty$  and  $p, q \rightarrow \infty$ .*

*Proof* The approximations  $\tilde{c}_{m,n}(s, t)$  and  $\sigma_{p,q}(s, t)$  of  $c(s, t)$  correspond to truncations of the generalized spectral density  $s(v, \eta)$  and of the covariance function  $c(s, t)$ , respectively. For arbitrary but fixed  $I(\bar{v}, \bar{\eta})$  and  $J(\bar{\tau}, \bar{\rho})$  we have

$$\begin{aligned}
|\tilde{c}_{m,n}(s, t) - \sigma_{p,q}(s, t)| &\leq |\tilde{c}_{m,n}(s, t) - \tilde{c}(s, t)| + |\tilde{c}(s, t) - c(s, t)| \\
&\quad + |c(s, t) - c(s, t) 1((s, t) \in J(\bar{\tau}, \bar{\rho}))| + |c(s, t) 1((s, t) \in J(\bar{\tau}, \bar{\rho})) - \sigma_{p,q}(s, t)|.
\end{aligned}$$

The first term can be bounded by

$$\begin{aligned}
|\tilde{c}_{m,n}(s, t) - \tilde{c}(s, t)| &\leq \sum_{k=-m}^m \sum_{l=-n}^n \int_{I_{k,l}} |s(v, \eta)| |e^{i(vs - \eta t)} - e^{i(v_k s - \eta_l t)}| dv d\eta \\
&\sim O(\max(\Delta v, \Delta \eta)) \sum_{k=-m}^m \sum_{l=-n}^n \int_{I_{k,l}} |s(v, \eta)| dv d\eta \\
&= O(\max(\Delta v, \Delta \eta)) \int_{I(\bar{v}, \bar{\eta})} |s(v, \eta)| dv d\eta,
\end{aligned}$$

so that  $|\tilde{c}_{m,n}(s, t) - \tilde{c}(s, t)| \rightarrow 0$  as  $m, n \rightarrow \infty$  since  $\Delta v = \bar{v}/m$ ,  $\Delta \eta = \bar{\eta}/n$ , and  $s(v, \eta)$  is assumed to be continuous and absolutely integrable. The convergence  $|c(s, t)1((s, t) \in J(\bar{\tau}, \bar{\rho})) - \sigma_{p,q}(s, t)| \rightarrow 0$  as  $p, q \rightarrow \infty$  for  $(s, t) \in J(\bar{\tau}, \bar{\rho})$  holds by Theorem 3.50, so that the fourth term can be made as small as desired. The terms  $|\tilde{c}(s, t) - c(s, t)|$  and  $|c(s, t) - c(s, t)1((s, t) \in J(\bar{\tau}, \bar{\rho}))|$  can be made arbitrarily small by increasing the truncation levels in the frequency and the time domains, that is, the sizes of the rectangles  $I(\bar{v}, \bar{\eta})$  and  $J(\bar{\tau}, \bar{\rho})$  (Theorem 3.51).  $\blacktriangle$

**Theorem 3.53** *If  $\bar{v}/m = \pi/\bar{\tau}$ ,  $\bar{\eta}/n = \pi/\bar{\rho}$ ,  $m = p$ ,  $n = q$ , and the conditions in Theorem 3.52 hold, the discrepancy between the coefficients  $\zeta_{k,l}$  and  $s_{k,l}$  of  $\sigma_{m,n}(s, t)$  in (3.90) and  $\tilde{c}_{m,n}(s, t)$  in (3.93) can be made arbitrarily small for sufficiently large truncation levels in both frequency and time domains as  $m, n \rightarrow \infty$  and  $p, q \rightarrow \infty$ .*

*Proof* If  $\bar{v}/m = \pi/\bar{\tau}$ ,  $\bar{\eta}/n = \pi/\bar{\rho}$ ,  $m = p$ , and  $n = q$ , then  $\sigma_{m,n}(s, t)$  and  $\tilde{c}_{m,n}(s, t)$  have the same frequencies, so that  $\sigma_{m,n}(s, t) - \tilde{c}_{m,n}(s, t) = \sum_{k=-m}^m \sum_{l=-n}^n (\zeta_{k,l} - s_{k,l}) e^{i(v_k s - \eta_l t)}$ . Since  $|\sigma_{m,n}(s, t) - \tilde{c}_{m,n}(s, t)|$  can be made arbitrarily small for sufficiently large truncation levels as  $m, n \rightarrow \infty$  and  $p, q \rightarrow \infty$  (Theorem 3.52), we can approximate  $\{\zeta_{k,l}\}$  by  $\{s_{k,l}\}$  so that the coefficients of  $\sigma_{m,n}(s, t)$  can be obtained from ordinates of the generalized spectral density  $s(v, \eta)$  of  $X(t)$ .

Note that the Fourier transform,

$$\begin{aligned} \tilde{s}_{m,n}(v, \eta) &= \sum_{k=-m}^m \sum_{l=-n}^n s_{k,l} \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{i(v_k s - \eta_l t)} e^{-i(v s - \eta t)} ds dt \\ &= \sum_{k=-m}^m \sum_{l=-n}^n s_{k,l} \delta(v - v_k) \delta(\eta - \eta_l), \end{aligned} \quad (3.94)$$

of  $\tilde{c}_{m,n}(s, t)$  in (3.93) is a generalized spectral density with power at frequencies  $(v_k, \eta_l)$ . The expression of  $\tilde{s}_{m,n}(v, \eta)$  results from the definition of the generalized spectral density in (3.85), the approximation  $\tilde{c}_{m,n}(s, t)$  of  $c(s, t)$ , and properties of the  $\delta$ -function.  $\blacktriangle$

Following is a Monte Carlo algorithm for generating samples of a real-valued non-stationary Gaussian process  $X(t)$  with mean 0, covariance function  $c(s, t) = E[X(s)X(t)]$ , and generalized spectral density  $s(v, \eta)$  in a bounded time interval  $[0, \tau]$  based on the sequence of processes

$$X^{(n)}(t) = \sum_{k=-n}^n C_k e^{i v_k t} = \frac{A_0}{2} + \sum_{k=1}^n \left( A_k \cos(v_k t) + B_k \sin(v_k t) \right), \quad n = 1, 2, \dots, \quad (3.95)$$

where  $v_k = k \Delta v$ ,  $\Delta v = \bar{v}/n$ ,  $\{C_k\}$  are complex-valued Gaussian variables with mean 0 and covariances  $E[C_k C_l^*] = s_{k,l}$ ,  $k, l = -n, \dots, n$ , and  $A_0 = 2C_0$ ,  $A_k = C_k + C_{-k}$ , and  $B_k = i(C_k - C_{-k})$ ,  $k = 1, \dots, n$ , are real-valued Gaussian variables. Note that  $X^{(n)}(t)$  are real-valued Gaussian processes,  $E[X^{(n)}(t)] = E[X(t)] = 0$  at all times  $t$ ,  $E[X^{(n)}(s)X^{(n)}(t)^*] = \sum_{k,l=-n}^n s_{k,l} \exp(i(v_k s - v_l t))$  converges to

$c(s, t)$  as the truncation level  $\bar{v}$  and the resolution  $n$  increase indefinitely (Theorems 3.50–3.53).

The generation of samples of  $X^{(n)}(t)$  in  $[0, \tau]$  involves the following three steps. First, if  $s(v, \eta)$  has unbounded support, select a bounded frequency range  $I(\bar{v}, \bar{v})$ ,  $0 < \bar{v} < \infty$ , such that  $\int_{I(\bar{v}, \bar{v})} s(v, \eta) e^{i(v-\eta)t} dv d\eta \simeq \int_{\mathbb{R}^2} s(v, \eta) e^{i(v-\eta)t} dv d\eta = c(t, t)$ . Otherwise, select  $I(\bar{v}, \bar{v})$  such that it includes the support of  $s(v, \eta)$ . Construct a partition  $\{I_{k,l}\}$  of  $I(\bar{v}, \bar{v})$  as defined by (3.91) with  $m = n$ , and impose the condition  $2\pi/\nu_1 = 2\pi/\Delta v = 2\pi n/\bar{v} \geq \tau$  to assure that the samples of  $X^{(n)}(t)$  are not periodic in  $[0, \tau]$ . Second, calculate the covariances of the real-valued dependent Gaussian random variables  $\{A_0, A_k, B_k, k = 1, \dots, n\}$  from those of  $\{C_k, k = -n, \dots, n\}$ , and generate samples of  $\{A_0, A_k, B_k, k = 1, \dots, n\}$  and corresponding samples of  $\{C_k, k = -n, \dots, n\}$ . Third, calculate samples of  $X^{(n)}(t)$ ,  $t \in [0, \tau]$ , from (3.95).

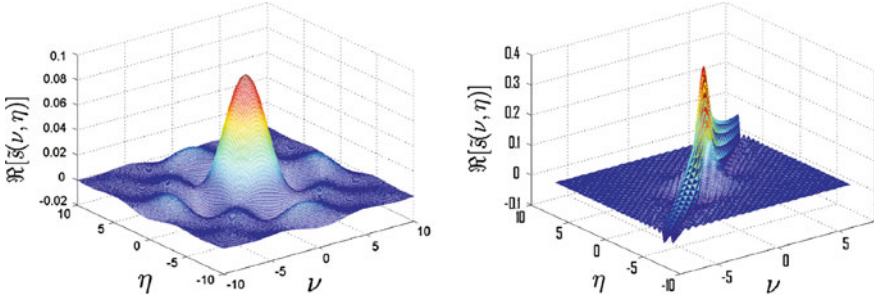
The algorithm is conceptually analogous to algorithms for generating samples of stationary Gaussian processes based on the spectral representation theorem (Sect. 3.8.1). Both algorithms require to approximate target spectral densities by discrete spectral densities with power at a finite number of frequencies and represent Gaussian processes by finite sums of harmonics with Gaussian coefficients. The only difference between these representations is that their coefficients are independent for stationary processes and dependent for nonstationary processes.

*Example 3.60* Let  $X(t) = U(h(t))$  be a real-valued process, where  $h: \mathbb{R} \rightarrow \mathbb{R}$  is a monotonically increasing, continuous function such that  $h(0) = 0$  and  $U(t)$  is a stationary Gaussian process with mean 0 and covariance function  $E[U(s)U(t)] = c_{st}(s - t) = (1 + \lambda|s - t|) \exp(-\lambda|s - t|)$ ,  $\lambda > 0$ . The process  $X(t)$  is Gaussian since the vectors  $(X(t_1), \dots, X(t_r))$  and  $(U(h(t_1)), \dots, U(h(t_r)))$  are equal in distribution for any integer  $r \geq 1$  and times  $t_1 < \dots < t_r$ , and  $(U(h(t_1)), \dots, U(h(t_r)))$  is a Gaussian vector. The mean and covariance functions of  $X(t)$  are  $E[X(t)] = E[U(h(t))] = 0$  and  $c(s, t) = E[X(s)X(t)] = E[U(h(s))U(h(t))] = c_{st}(h(s) - h(t))$ , so that  $X(t)$  is a nonstationary Gaussian process.

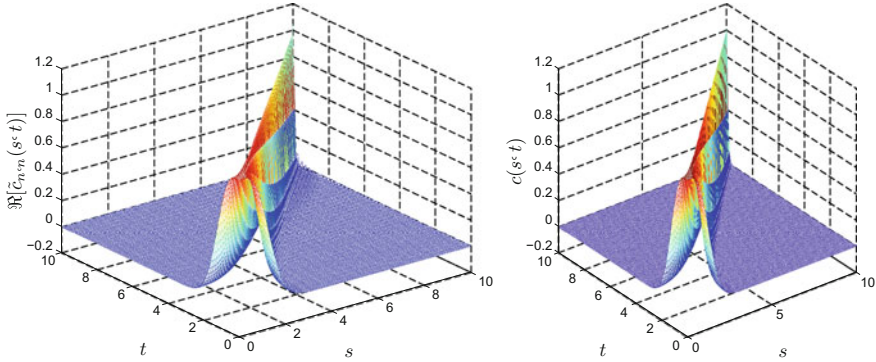
Numerical results in Figs. 3.15–3.17 are for  $h(t) = t^2 \text{sign}(t)$  and  $\lambda = 1$ . Figure 3.15 shows the real part of the approximate spectral densities  $\tilde{s}(v, \eta)$  in (3.87) for  $\bar{\tau} = \bar{\rho} = 1$  (left panel), and  $\bar{\tau} = \bar{\rho} = 10$  (right panel); the accuracy of the calculated generalized spectral density depends strongly on the size of the integration domain  $J(\bar{\tau}, \bar{\rho}) = [-\bar{\tau}, \bar{\tau}] \times [-\bar{\rho}, \bar{\rho}]$ . Integration domains  $J(\bar{\tau}, \bar{\rho})$  with  $\bar{\tau} = \bar{\rho} \geq 10$  yield insignificant changes in the real part of  $\tilde{s}(v, \eta)$ . For all values of  $\bar{\tau} = \bar{\rho}$ , the magnitude of the ordinates of the imaginary part of  $\tilde{s}(v, \eta)$  are of order  $10^{-16}$ . The right and left panels in Figs. 3.16 and 3.17 show the covariance function

$$c(s, t) = (1 + \lambda|s^2 \text{sign}(s) - t^2 \text{sign}(t)|) \exp(-\lambda|s^2 \text{sign}(s) - t^2 \text{sign}(t)|) \quad (3.96)$$

of  $X(t)$  and approximations of this covariance function. The left panel in Fig. 3.16 is  $\tilde{c}_{n,n}(s, t)$  in (3.93) with  $m = n = 400$  and  $\bar{v} = \bar{\eta} = 30$ . The left panel in Fig. 3.17 is an estimates of  $c(s, t)$  obtained from  $n_s = 1000$  independent samples of  $X^{(n)}(t)$  in (3.95) with  $n = 400$  generated in  $[0, \tau]$  with  $\tau = 10$ . The approximate covariance function and the estimate of this function are satisfactory. Further improvements can



**Fig. 3.15** Real parts of  $\tilde{s}(\nu, \eta)$  for  $\lambda = 1$ ,  $\bar{\tau} = 1$  (left panel), and  $\bar{\tau} = 10$  (right panel)



**Fig. 3.16** Approximate (left panel) and exact (right panel) covariance functions for  $\lambda = 1$ ,  $\bar{v} = \bar{\eta} = 30$ , and  $n = 400$

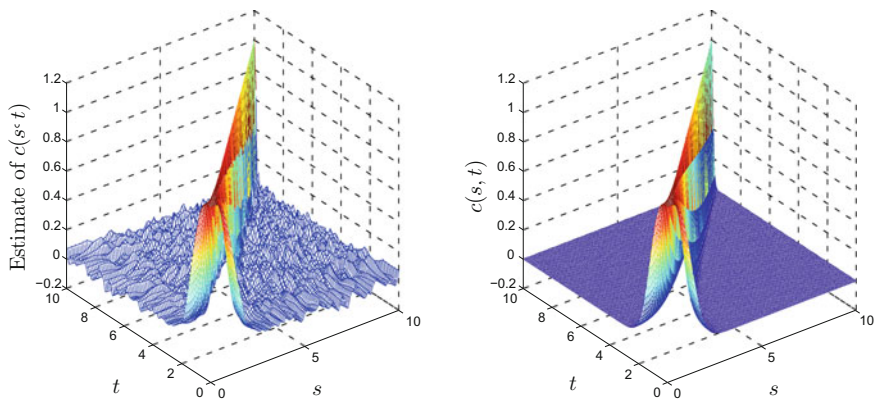
be obtained by increasing the resolution of the model  $X^{(n)}(t)$  of  $X(t)$ , which can be achieved by using larger  $\bar{\tau} = \bar{\eta}$  and/or  $m = n$ .  $\diamond$

Additional information on the spectral-based Monte Carlo algorithm for generating samples of non-stationary Gaussian processes discussed in this section can be found in [32]. In this reference, it also shown that the algorithm is sufficiently stable to be applied even to stationary processes, a severe test since the generalized spectral density for a stationary process  $X(t)$  with spectral density  $s_{\text{st}}(\nu)$  has the functional form  $s(\nu, \eta) = s_{\text{st}}(\nu)\delta(\nu - \eta)$ ,  $(\nu, \eta) \in \mathbb{R}^2$ .

### 3.9 Exercises

**Exercise 3.1** Show that the correlation function of an  $\mathbb{C}^d$ -valued random function  $X$  defined on  $\mathbb{R}^{d'}$  has the properties  $r_{i,j}(s, t) = r_{j,i}(t, s)^*$ ,  $|r_{i,j}(s, t)|^2 \leq r_{i,i}(s, s)r_{j,j}(t, t)$ , and  $\sum_{k,l=1}^n r_{i,j}(t_k, t_l)z_k z_l^*$  is real and positive for arbitrary  $n \geq 1$ ,  $t_k \in \mathbb{R}^{d'}$ , and  $z_k \in \mathbb{C}$ .





**Fig. 3.17** Estimated (*left panel*) and exact (*right panel*) covariance functions for  $\lambda = 1$ ,  $\bar{\nu} = \bar{\eta} = 30$

**Exercise 3.2** Show that the process  $X(t)$  in Example 3.14 is Gaussian if  $\{A_k\}$  and  $\{B_k\}$  are independent Gaussian variables. Find the finite dimensional distributions of  $X(t)$  under the assumptions that  $\{A_k\}$  and  $\{B_k\}$  are independent and dependent Gaussian variables.

**Exercise 3.3** Suppose the argument  $t$  in Example 3.13 is a spatial coordinate  $t \in \mathbb{R}^{d'}$ , so that  $X$  is a real-valued random field. Show that  $X$  is weakly homogeneous with mean  $E[X(t)] = 0$  and covariance function  $c(s, t) = \sum_{k=1}^n \sigma_k^2 \cos(v_k \cdot (s - t))$ , where  $v_k \in \mathbb{R}^{d'}$  and  $\sigma_k > 0$  are constants.

**Exercise 3.4** Show that the correlation function  $r(\tau) = E[X(t + \tau)X(t)^*]$  of a complex-valued weakly stationary process  $X$  satisfies the condition  $r(\tau) = r(-\tau)^*$ .

**Exercise 3.5** Let  $X$  be a real-valued weakly homogeneous random field defined on  $\mathbb{R}^{d'}$ . Show the spectral distribution  $S(v)$  of this field given by (3.17) satisfies the condition  $\int_{\mathbb{R}^{d'}} \sin(t \cdot v) dS(v) = 0$ .

**Exercise 3.6** Construct examples of m.s. continuous and m.s. differentiable stochastic processes and random fields.

**Exercise 3.7** Complete the proof of Theorem 3.13.

**Exercise 3.8** Show that  $Y$  in Example 3.29 satisfies the differential equation  $\dot{Y}(t) = g(t)Y(t)$ ,  $t \geq 0$ , with initial condition  $Y(0) = 0$ , if  $h$  is differentiable and  $g(t) = dh(t)/dt$ .

**Exercise 3.9** Show that  $X(t) = \int_{\mathbb{R}^{d'}} [\cos(t'v) dU(v) + \sin(t'v) dV(v)]$  is an alternative representation for the random field  $X$  in Theorem 3.20, where  $U$  and  $V$  are real-valued random fields satisfying the conditions  $E[dU(v)] = E[dV(v)] = 0$ ,  $E[dU(v)dV(v')] = 0$ , and  $E[dU(v)dU(v')] = E[dV(v)dV(v')] = dS(v)\delta(v - v')$ .

**Exercise 3.10** Show that the random variables  $\{X_k\}$  in the Karhunen–Loève expansion of a random function satisfying the conditions of Theorem 3.22 have the properties given by (3.32) and  $E[(X(s) - X^{(n)}(s))(X(t) - X^{(n)}(t))^*] = r(s, t) - \sum_{k=1}^n \lambda_k \phi_k(s) \phi_k(t)^* \rightarrow 0$  as  $n \rightarrow \infty$ , where  $X^{(n)}(t) = \sum_{k=1}^n \lambda_k^{1/2} X_k \phi_k(t)$ .

**Exercise 3.11** Show that a weakly stationary Gaussian function is also stationary and that linear transformations of Gaussian functions are Gaussian.

**Exercise 3.12** Calculate the scaled covariance functions in Examples 3.38 and 3.39.

**Exercise 3.13** Let  $\{X_n, n = 0, 1, \dots\}$  be a Markov chain taking values in  $S = \{1, 2, \dots, q\}$ . Prove the relationships in (3.41).

**Exercise 3.14** Prove the Chapman–Kolmogorov equation in Theorem 3.30 by using properties of conditional densities and Markov processes.

**Exercise 3.15** Show that the expectation of both discrete and continuous time submartingales, martingales, and supermartingales are increasing, constant, and decreasing functions of time.

**Exercise 3.16** Find the characteristic functions for Poisson and compound Poisson processes.

**Exercise 3.17** Show that the period between consecutive jumps of a Poisson process  $N$  with intensity  $\lambda$  is an exponential random variable with mean  $1/\lambda$ .

**Exercise 3.18** Find the quadratic variation process for  $B + \tilde{C}$ , where  $B$  is Brownian motion,  $\tilde{C}(t) = C(t) - \lambda t E[Y_1]$ ,  $C(t) = \sum_{k=1}^{N(t)} Y_k$ ,  $\{Y_k\}$  are iid random variables with finite mean, and  $N(t)$  denotes a Poisson process with intensity  $\lambda > 0$ . Assume that  $B$  and  $C$  are independent of each other.

**Exercise 3.19** Find the quadratic variation process for the approximate representation  $\tilde{L}_{\alpha,a}(t)$  given by (3.62) for an  $\alpha$ -stable process  $L_\alpha(t)$ .

## References

1. Samorodnitsky G, Taqqu MS, (1994) Stable non-Gaussian random processes. Stochastic models with infinite variance. Birkhäuser, New York
2. Adler RJ (1981) The geometry of random fields. Wiley, New York
3. Applebaum D (2004) Lévy processes and stochastic calculus. Cambridge University Press, Cambridge
4. Grigoriu M (1995) Applied non-Gaussian processes: Examples, theory, simulation, linear random vibration, and MATLAB solutions. Prentice Hall, Englewoods Cliffs
5. Grigoriu M (2002) Stochastic calculus. Applications in science and engineering. Birkhäuser, Boston
6. Protter P (1990) Stochastic integration and differential equations. Springer, New York
7. Wong E, Hajek B (1985) Stochastic processes in engineering systems. Springer, New York
8. Cramer H, Leadbetter MR (1967) Stationary and related stochastic processes. Wiley, New York
9. Matérn B (1986) Spatial variation, 2nd edn. Springer, New York

10. Soong TT, Grigoriu M (1993) Random vibration of mechanical and structural systems. Prentice Hall, Englewood Cliffs
11. Ruymgaart PA, Soong TT (1988) Mathematics of Kalman–Bucy filtering. Springer, New York
12. Davenport WB, Root WL (1958) An introduction to the theory of random signals and noise. McGraw-Hill Book Company, New York
13. Hernández DB (1995) Lectures on probability and second order random fields. World Scientific, London
14. Grigoriu M (2006) Evaluation of Karhunen–Loève, spectral, and sampling representations for stochastic processes. *J Eng Mech ASCE*, 132(2):179–189
15. Lancaster P, Tismenetsky M (1985) The theory of matrices, 2nd edn. Academic Press, New York
16. Grigoriu M (2009) Existence and construction of translation models for stationary non-Gaussian processes. *Probab Eng Mech* 24:545–551
17. Meyn SP, Tweedie RL (1993) Markov chains and stochastic stability. Springer, New York
18. Grigoriu M (2010) Nearest neighbor probabilistic model for aluminum polycrystals. *J Eng Mech* 136(7):821–829
19. Tjøstheim D (1978) Statistical spatial series modelling. *Adv Appl Probab* 10:130–154
20. Tjøstheim D (1981) Autoregressive modeling and spectral analysis of array data in the plane. *IEEE Trans Geosci Remote Sens* GE19(1):15–24
21. Whittle P (1954) On stationary processes in the plane. *Biometrika* 41(3/4):434–449
22. Çinlar E (1975) Introduction to stochastic processes. Prentice Hall, Englewood Cliffs
23. Resnick SI (1992) Adventures in stochastic processes. Birkhäuser, Boston
24. Ethier SN, Kurtz TG (1986) Markov processes. Characterization and convergence. Wiley, New York
25. Steele JM (2001) Stochastic calculus and financial applications. Springer, New York
26. Snyder DL (1975) Random point processes. Wiley, New York
27. Asmussen S, Rosiński J (2001) Approximations of small jumps of Lévy processes with a view towards simulation. *J Appl Probab* 38:482–493
28. Ogorodnikov VA, Prigarin SM (1996) Numerical modelling of random processes and fields: Algorithms and applications. VSP BV, Utrecht
29. Prigarin SM (2001) Spectral models of random fields in Monte Carlo simulation. VSP BV, Boston
30. Papoulis A (1965) Probability, random variables, and stochastic processes. McGraw-Hill Book Company, New York
31. Tolstov GP (1962) Fourier series. Dover Publications, New York
32. Grigoriu M (2010) A spectral-based Monte Carlo algorithm for generating samples of nonstationary Gaussian processes. *Monte Carlo Methods Appl* 16(2):143–165

# Chapter 4

## Stochastic Integrals

### 4.1 Introduction

The states of most physical systems are random functions that can be defined as solutions of differential equations driven by uncertain actions. For example, stochastic processes can be used to characterize vibrations in buildings induced by wind and stresses in aircrafts since wind loads and flight conditions fluctuate randomly in space and time.

Let  $X(t)$ ,  $t \geq 0$ , be an  $\mathbb{R}^d$ -valued process representing the state of a dynamic system in random environment. The process can be defined by a differential equation of the type

$$dX(t) = a(X(t-)) dt + b(X(t-)) dY(t), \quad t \geq 0, \quad (4.1)$$

where  $a(\cdot)$  and  $b(\cdot)$  are  $(d, 1)$  and  $(d, d')$  matrices and  $Y(t)$  is an  $\mathbb{R}^{d'}$ -valued input process. The meaning of (4.1) is given by its integral form

$$X(t) = X(0) + \int_0^t a(X(s-)) ds + \int_0^t b(X(s-)) dY(s), \quad t \geq 0. \quad (4.2)$$

The use of the left limit  $X(t-) = \lim_{u \uparrow t} X(u)$  in (4.2) will be clarified in a subsequent section. Conditions for the existence and uniqueness of the solution of (4.1) and (4.2) are discussed in Sect. 5.5.1. The integral  $\int_0^t b(X(s-)) dY(s)$  in (4.2) has random integrand and integrator, and the formal derivative of the integrator is usually a white noise process. We will see that this integrals is not, generally, defined in the Riemann-Stieltjes sense.

It is common in applications to assume that the formal derivative of  $Y(t)$  is a white noise process, for example, a Gaussian white noise if  $Y(t)$  is an  $\mathbb{R}^{d'}$ -valued Brownian motion  $B(t)$ . The assumption of white input is not restrictive since many colored noise processes can be defined as outputs of linear filters to white noise. For example, let  $U(t)$  denote the displacements of a simple oscillator with damping ratio  $\zeta > 0$  and natural frequency  $\nu_0$  subjected to a stationary first order Gauss Markov process

$V(t)$  with mean 0, variance 1, and correlation function  $E[V(s)V(t)] = \exp(-\lambda|s-t|)$ ,  $\lambda > 0$ , so that  $V(t)$  is the solution of  $dV(t) = -\lambda V(t)dt + \sqrt{2\lambda}dB(t)$  with initial state  $V(0) \sim N(0, 1)$  assumed to be independent of the Brownian motion  $B(t)$ . The equation of motion for the oscillator,  $\ddot{U}(t) + 2\zeta\nu_0\dot{U}(t) + \nu_0^2 U(t) = V(t)$ , augmented with that for the input satisfies the differential equation

$$dX(t) = \begin{bmatrix} X_2(t) \\ -\nu_0^2 X_1(t) - 2\zeta\nu_0 X_2(t) + X_3(t) \\ -\lambda X_3(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ 0 \\ \sqrt{2\lambda} \end{bmatrix} dB(t) \quad (4.3)$$

with the notation  $X(t) = (X_1(t) = U(t), X_2(t) = \dot{U}(t), X_3(t) = V(t))$ , that is, an equation of the type in (4.1).

## 4.2 Riemann-Stieltjes Integrals

Let  $f$  and  $g$  be real-valued functions defined on a bounded interval  $[0, t] \subset \mathbb{R}$ ,  $p = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 < \dots < t_n = t$ , be a partition of  $[0, t]$ , and  $t'_k \in [t_{k-1}, t_k]$  denote intermediate points of  $p$ . Let

$$s_{f,p}(g) = \sum_{k=1}^n f(t'_k) [g(t_k) - g(t_{k-1})] \quad (4.4)$$

be the Riemann-Stieltjes sum of  $f$  with respect to  $g$  for the partition  $p$  of  $[0, t]$ . The function  $f$  is Riemann-Stieltjes integrable with respect to  $g$  on  $[0, t]$  if there is a number  $a$  such that, for every  $\varepsilon > 0$ , there is a partition  $p_\varepsilon$  of  $[0, t]$  with the property  $|s_{f,p}(g) - a| < \varepsilon$  for every  $p \supseteq p_\varepsilon$  and every choice of intermediate points ([1], p. 141). The number  $a$  is called the Riemann-Stieltjes integral of  $f$  with respect to  $g$  on  $[0, t]$  and is denoted by  $\int_0^t f(s) dg(s)$  or  $\int_0^t f dg$ . It can be shown that  $\int_0^t f dg$  exists if (1)  $f$  and  $g$  have no discontinuities at the same point  $s \in [0, t]$  and (2)  $f$  and  $g$  are of bounded  $p$ -variation and  $q$ -variation, respectively, for some  $p > 0$  and  $q > 0$  such that  $1/p + 1/q > 1$  ([5], p. 94). A real-valued function  $h$  is of bounded  $r$ -variation on  $[0, t]$  if  $\sup \left\{ \sum_k |h(t_k) - h(t_{k-1})|^r \right\} < \infty$ , where  $r > 0$  and the supremum is taken over all partitions of the interval  $[0, t]$ . If  $r = 1$ ,  $h$  is said to be of bounded variation on  $[0, t]$ . Real-valued differentiable functions on a bounded interval of the real line are of bounded variation. However, continuous functions may not have this property. For example, the continuous function  $h : [0, 1] \rightarrow \mathbb{R}$  defined by  $h(x) = 0$  if  $x = 0$  and  $h(x) = x \cos(\pi/x)$  if  $x \neq 0$  is not of bounded variation.

The Riemann-Stieltjes integral has the properties

$$\begin{aligned} \int_0^t (c_1 f_1 + c_2 f_2) dg &= c_1 \int_0^t f_1 dg + c_2 \int_0^t f_2 dg, \\ \int_0^t f d(c_1 g_1 + c_2 g_2) &= c_1 \int_0^t f dg_1 + c_2 \int_0^t f dg_2, \quad \text{and} \\ \int_0^t f dg &= \int_0^s f dg + \int_s^t f dg, \end{aligned} \quad (4.5)$$

provided  $f_1, f_2$  are Riemann-Stieltjes integrable with respect to  $g$  on  $[0, t]$  and  $f$  is Riemann-Stieltjes integrable with respect to  $g, g_1, g_2$  on  $[0, t]$ , where  $c_1, c_2$  are constants and  $s \in (0, t)$  ([1], Theorems 7.2, 7.3, and 7.4).

It is natural to attempt a construction of the Riemann-Stieltjes integral for the more general case in which the integrand and the integrator are random functions. Suppose first that the integrator is a Brownian motion  $B$ . The samples  $s \mapsto B(s, \omega)$  of  $B$  are of bounded  $q$ -variation on any finite interval  $[0, t]$  for  $q > 2$  [8]. Hence,  $\int_0^t f(s) dB(s, \omega)$  exists as a Riemann-Stieltjes integral for almost all sample paths of  $B$  if  $f$  is of bounded variation since  $1/p + 1/q > 1$  for  $p = 1$  and  $q > 2$ . The definition of  $\int_0^t f dB$  as a Riemann-Stieltjes integral corresponding to the sample paths of  $B$  is referred to as path-by-path definition. The integrals

$$\int_0^t e^s dB(s, \omega), \quad \int_0^t \cos(s) dB(s, \omega), \quad \text{and} \quad \int_0^t s^k dB(s, \omega)$$

exist as Riemann-Stieltjes integrals for almost all samples of the Brownian motion because the functions  $e^s$ ,  $\cos(s)$ , and  $s^k$  are of bounded variation.

Consider now more general integrals, for example,  $\int_0^t B dB$  and  $\int_0^t f dB$ , where  $f$  is an arbitrary continuous function. The path-by-path Riemann-Stieltjes integral  $\int_0^t B(s, \omega) dB(s, \omega)$  does not exist since the samples of  $B$  are of bounded  $q$ -variation for  $q > 2$  so that the condition  $1/p + 1/q > 1$  is not satisfied. The Riemann-Stieltjes integral  $\int_0^t f(s) dB(s, \omega)$  cannot be defined as a path by path Riemann-Stieltjes integral for  $f$  continuous since  $\int_0^t f(s) dg(s)$  does not exist for all continuous functions  $f$  on  $[0, t]$  unless  $g$  is of bounded variation ([6], Theorem 52, p. 40). We need an alternative definition for  $\int_0^t B dB$ ,  $\int_0^t f dB$ , and other more general integrals with random integrands and integrators.

### 4.3 Stochastic Integrals $\int B dB$ and $\int N dN$

Our objective is to extend the Riemann-Stieltjes integral to stochastic integrals of the type  $\int_I X dY$ , where  $X$  and  $Y$  are real-valued stochastic processes and  $I$  is an interval of the real line.

Stochastic integrals are defined by limits of sequences of sums resembling the Riemann-Stieltjes sums in (4.4). There are two significant differences between the

definition of stochastic integrals and Riemann-Stieltjes integrals. First, the sequence of sums defining stochastic integrals are random so that convergence criteria for sequences of random variables have to be used to construct stochastic integrals. Second, the limit of the sequence of sums defining stochastic integrals may depend on the selection of the intermediate points  $\{t'_k\}$ , in contrast to Riemann-Stieltjes sums whose limits are independent of the particular selection of these points.

The stochastic integrals  $\int_0^t B dB$  and  $\int_0^t N dN$  are used to illustrate both the construction of stochastic integrals and differences between stochastic and Riemann-Stieltjes integrals, where  $B$  and  $N$  denote the Brownian motion and Poisson processes, respectively. Two stochastic integrals are defined for particular selections of the intermediate points, the Itô and the Stratonovich integrals. It is shown that (1) the Itô integral of the Brownian motion with respect to itself is a martingale while the corresponding Stratonovich integral is not and (2) the Itô and the path-wise Riemann-Stieltjes definitions of the integral  $\int_0^t N(s-) dN(s)$  coincide, where  $N(s-) = \lim_{u \uparrow s} N(u)$ .

*Example 4.1* Consider the random sequence

$$J_{B,n}(B) = \sum_{k=1}^n B(t_{k-1}) [B(t_k) - B(t_{k-1})],$$

where  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 < t_1 < \dots < t_n = t$ , is a sequence of partitions of  $[0, t]$  with intermediate points  $t'_k = t_{k-1}$  such that  $\Delta(p_n) = \max_{1 \leq k \leq n} (t_k - t_{k-1}) \rightarrow 0$  as  $n \rightarrow \infty$ . The limit of  $J_{B,n}(B)$  as  $n \rightarrow \infty$  exists in m.s. and in probability. It is denoted by  $\int_0^t B(s) dB(s)$  or  $\int_0^t B dB$ , has the expression

$$\int_0^t B(s) dB(s) = \frac{1}{2}(B(t)^2 - t), \quad t \geq 0, \quad (4.6)$$

and is called the Itô or the stochastic integral of  $B$  with respect to  $B$  on  $[0, t]$ . If, in addition, the sequence of partitions of  $[0, t]$  is refining, that is,  $p_n \subset p_{n+1}$ , then  $J_{B,n}(B)$  converges a.s. to the limit in (4.6). The Itô integral differs from  $B(t)^2/2$ , that is, the integral  $\int_0^t B dB$  delivered by the formal use of classical calculus.  $\diamond$

*Proof* The random variable  $J_{B,n}(B)$  can be given in the form

$$\begin{aligned} J_{B,n}(B) &= \frac{1}{2} \sum_{k=1}^n (B(t_k)^2 - B(t_{k-1})^2) - \frac{1}{2} \sum_{k=1}^n (B(t_k) - B(t_{k-1}))^2 \\ &= \frac{1}{2} B(t)^2 - \frac{1}{2} \sum_{k=1}^n (B(t_k) - B(t_{k-1}))^2, \end{aligned}$$

where the last equality holds since  $\sum_{k=1}^n (B(t_k)^2 - B(t_{k-1})^2)$  is a telescopic series whose sum is  $B(t)^2$ . The m.s. limit of  $J_{B,n}(B)$  is  $(B(t)^2 - t)/2$  since

$$\begin{aligned}
E\left[\left(J_{B,n}(B) - \frac{1}{2}(B(t)^2 - t)\right)^2\right] &= E\left[\left(-\frac{1}{2}\sum_{k=1}^n \left(B(t_k) - B(t_{k-1})\right)^2 + \frac{t}{2}\right)^2\right] \\
&= \frac{1}{4}\sum_{k,l} E[(\Delta B_k)^2(\Delta B_l)^2] - \frac{t}{2}\sum_k E[(\Delta B_k)^2] + \frac{t^2}{4} \rightarrow 0, \quad \text{as } n \rightarrow \infty,
\end{aligned}$$

where  $\Delta B_k = B(t_k) - B(t_{k-1})$  (Exercise 4.1). Hence, the Itô integral  $\int_0^t B dB$  can be defined as the m.s. limit of  $J_{B,n}(B)$ , and as the limit in probability of  $J_{B,n}(B)$  since convergence in probability is implied by m.s. convergence.

Since  $\lim_{n \rightarrow \infty} \sum_{k=1}^n [B(t_k) - B(t_{k-1})]^2 = t$  holds a.s. for a sequence of refining partitions of  $[0, t]$  [6] (Theorem 28, p. 18),  $J_{B,n}(B)$  converges a.s. to  $(B(t)^2 - t)/2$  for refining partitions so that the Itô integral  $\int_0^t B dB$  can be defined as the a.s. limit of  $J_{B,n}(B)$  for these types of partitions. ▲

*Example 4.2* Consider the random sequence

$$\tilde{J}_{B,n}(B) = \sum_{k=1}^n B(t'_k)[B(t_k) - B(t_{k-1})],$$

where  $t'_k = (1 - \theta)t_{k-1} + \theta t_k$ ,  $\theta \in [0, 1]$ .  $\tilde{J}_{B,n}(B)$  differs from  $J_{B,n}(B)$  in Example 4.1 by the choice of the intermediate points, that is,  $t'_k$  instead of  $t_{k-1}$ . The limit

$$\lim_{n \rightarrow \infty} \tilde{J}_{B,n}(B) = \frac{1}{2} B(t)^2 + (\theta - 1/2) t. \quad (4.7)$$

exists in m.s. and probability. For  $\theta = 1/2$ , this limit, denoted by  $\int_0^t B(s) \circ dB(s)$  or  $\int_0^t B \circ dB$  and called the Stratonovich integral of  $B$  with respect to  $B$  on  $[0, t]$ , is

$$\int_0^t B(s) \circ dB(s) = \int_0^t B \circ dB = \frac{1}{2} B(t)^2, \quad (4.8)$$

and coincides with the result obtained by the formal use of classical calculus. ◇

*Proof* An alternative form of  $\tilde{J}_{B,n}(B)$  is

$$\tilde{J}_{B,n}(B) = \sum_{k=1}^n B(t_{k-1}) \Delta B_k + \sum_{k=1}^n \Delta B'_k \Delta B_k$$

with the notations  $\Delta t_k = t_k - t_{k-1}$ ,

$$\begin{aligned}
\Delta B_k &= B(t_k) - B(t_{k-1}) \sim N(0, \Delta t_k), \\
\Delta B'_k &= B(t'_k) - B(t_{k-1}) \sim N(0, \theta \Delta t_k), \quad \text{and} \\
\Delta B''_k &= B(t_k) - B(t'_k) \sim N(0, (1 - \theta) \Delta t_k)
\end{aligned}$$

for a fixed  $\theta \in [0, 1]$ . We have seen that the first term of  $\tilde{J}_{B,n}(B)$  converges in m.s. to  $(B(t)^2 - t)/2$  as  $n \rightarrow \infty$  and its limit defines the Itô integral in (4.6). It remains



to show that the second term in  $\tilde{J}_{B,n}(B)$ , that is,  $\sum_{k=1}^n \Delta B'_k \Delta B_k$ , converges in m.s. to  $\theta t$  since  $X_n + Y_n \xrightarrow{\text{m.s.}} X + Y$  holds if  $X_n \xrightarrow{\text{m.s.}} X$  and  $Y_n \xrightarrow{\text{m.s.}} Y$ . The first moment of the second term of  $\tilde{J}_{B,n}(B)$  is

$$E \left[ \sum_{k=1}^n \Delta B'_k \Delta B_k \right] = \sum_k E[(\Delta B'_k)^2] + \sum_k E[\Delta B'_k \Delta B''_k] = \sum_k \theta \Delta t_k = \theta t$$

since  $\Delta B_k = \Delta B'_k + \Delta B''_k$  and  $E[\Delta B'_k \Delta B''_k] = E[\Delta B'_k]E[\Delta B''_k] = 0$ . The second moment of this term,

$$\begin{aligned} E \left[ \left( \sum_{k=1}^n \Delta B'_k \Delta B_k \right)^2 \right] &= E \left[ \left( \sum_{k=1}^n (\Delta B'_k)^2 + \sum_{k=1}^n \Delta B'_k \Delta B''_k \right)^2 \right] \\ &= \sum_{k \neq l} E[(\Delta B'_k)^2 (\Delta B'_l)^2] + \sum_k E[(\Delta B'_k)^4] + \sum_{k,l} E[\Delta B'_k \Delta B''_k \Delta B'_l \Delta B''_l] \\ &\quad + 2 \sum_{k,l} E[(\Delta B'_k)^2 \Delta B'_l \Delta B''_l] = \theta^2 \sum_{k \neq l} \Delta t_k \Delta t_l + 3 \theta^2 \sum_k (\Delta t_k)^2 \\ &\quad + \theta(1-\theta) \sum_k (\Delta t_k)^2 = \theta^2 \sum_{k,l} \Delta t_k \Delta t_l + 2 \theta^2 \sum_k (\Delta t_k)^2 + \theta(1-\theta) \sum_k (\Delta t_k)^2, \end{aligned}$$

converges to  $(\theta t)^2$  as  $n \rightarrow \infty$  since  $\sum_{k,l} \Delta t_k \Delta t_l = t^2$  and  $\sum_k (\Delta t_k)^2 \leq \max_l \{\Delta t_l\} \sum_k \Delta t_k$ ,  $\max_l \{\Delta t_l\} \rightarrow 0$  as  $\Delta(p_n) \rightarrow 0$ , and  $\sum_k \Delta t_k = t$ . Hence,  $\sum_{k=1}^{m_n} \Delta B'_k \Delta B_k$  converges in m.s. to  $\theta t$  so that the m.s. limit of  $\tilde{J}_{B,n}(B)$  exists and is

$$\lim_{n \rightarrow \infty} \tilde{J}_{B,n}(B) = \frac{1}{2} (B(t)^2 - t) + \theta t = \frac{1}{2} B(t)^2 + (\theta - 1/2) t.$$

This limit coincides with the Itô integral for  $\theta = 0$  but differs from this integral for  $\theta \neq 0$ . The Stratonovich integral corresponds to  $\theta = 1/2$ .  $\blacktriangle$

*Example 4.3* Denote by

$$\begin{aligned} J_{B,n}(B)(s) &= \sum_{k=1}^n B(t_{k-1}) [B(t_k \wedge s) - B(t_{k-1} \wedge s)] \quad \text{and} \\ \tilde{J}_{B,n}(B)(s) &= \sum_{k=1}^n B(t'_k) [B(t_k \wedge s) - B(t_{k-1} \wedge s)] \end{aligned}$$

the restrictions of  $J_{B,n}(B)$  and  $\tilde{J}_{B,n}(B)$  to intervals  $[0, s]$ ,  $s \leq t$ . Then  $J_{B,n}(B)(s)$  is an  $\mathcal{F}_s$ -square integrable martingale while  $\tilde{J}_{B,n}(B)(s)$  is not a martingale, where  $\mathcal{F}_s = \sigma(B(u) : 0 \leq u \leq s)$  and  $t'_k = (t_{k-1} + t_k)/2$ .  $\diamond$

*Proof* That  $J_{B,n}(B)(s)$  and  $\tilde{J}_{B,n}(B)(s)$  have finite second moments follows from Examples 4.1 and 4.2. Let  $k(s)$  be an index such that  $s \in [t_{k(s)-1}, t_{k(s)}]$ . Since  $J_{B,n}(B)(s) = \sum_{k=1}^{k(s)-1} B(t_{k-1}) [B(t_k) - B(t_{k-1})] + B(t_{k(s)-1}) [B(s) - B(t_{k(s)-1})]$

depends on  $B(u)$ ,  $u \leq s$ , it is  $\mathcal{F}_s = \sigma(B(u), 0 \leq u \leq s)$ -adapted. For  $0 \leq s \leq \sigma \leq t$  such that  $s < \sigma$  are in the same interval  $[t_{m-1}, t_m]$  of a partition  $p_n$  of  $[0, t]$ , we have  $J_{B,n}(B)(\sigma) = J_{B,n}(B)(s) + B(t_{m-1})(B(\sigma) - B(s))$  so that  $E[J_{B,n}(B)(\sigma) | \mathcal{F}_s] = J_{B,n}(B)(s) + B(t_{m-1})E[B(\sigma) - B(s) | \mathcal{F}_s] = J_{B,n}(B)(s)$  by properties of Brownian motion and conditional expectation. For  $0 \leq s \leq \sigma \leq t$  such that  $s$  and  $\sigma$  belong to distinct intervals of  $p_n$  of  $[0, t]$ , we have

$$\begin{aligned} J_{B,n}(B)(\sigma) - J_{B,n}(B)(s) &= B(t_{k(s)-1})(B(t_{k(s)}) - B(s)) \\ &\quad + \sum_{k=k(s)+1}^n B(t_{k-1})(B(t_k \wedge \sigma) - B(t_{k-1} \wedge \sigma)) \end{aligned}$$

so that  $E[J_{B,n}(B)(\sigma) - J_{B,n}(B)(s) | \mathcal{F}_s] = 0$  since  $E[B(t_{k(s)-1})(B(t_{k(s)}) - B(s)) | \mathcal{F}_s] = B(t_{k(s)-1})E[B(t_{k(s)}) - B(s) | \mathcal{F}_s] = 0$  and  $E[\sum_{k=k(s)+1}^n B(t_{k-1})\Delta B_{k,\sigma} | \mathcal{F}_s] = \sum_{k=k(s)+1}^n E\{B(t_{k-1})E[\Delta B_{k,\sigma} | \mathcal{F}_{t_{k-1}}] | \mathcal{F}_s\} = 0$  since  $E[\Delta B_{k,\sigma} | \mathcal{F}_{t_{k-1}}] = 0$ , where  $\Delta B_{k,\sigma} = B(t_k \wedge \sigma) - B(t_{k-1} \wedge \sigma)$ . Hence,  $J_{B,n}(B)(s)$  has the martingale property  $E[J_{B,n}(B)(\sigma) | \mathcal{F}_s] = J_{B,n}(B)(s)$  for any  $\sigma \geq s$ . That  $J_{B,n}(B)$  is a martingale follows from Theorem 4.4 discussed in Sect. 4.4.2. The process  $\tilde{J}_{B,n}(B)(s)$  is not adapted with respect to the filtration generated by  $B$ , for example,  $\tilde{J}_{B,n}(B)(s)$  with  $\theta = 1/2$  depends on  $B(u)$ ,  $u > s$ , for  $s \in [t_{k-1}, (t_{k-1} + t_k)/2]$ , so that it is not a martingale.  $\blacktriangle$

*Example 4.4* Consider the random sequence

$$J_{N,n}(N) = \sum_{k=1}^n N(t_{k-1})[N(t_k) - N(t_{k-1})],$$

where  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 < t_1 < \dots < t_n = t$ , is a sequence of partitions of  $[0, t]$  such that  $\Delta(p_n) \rightarrow 0$  as  $n \rightarrow \infty$  and  $N$  is a Poisson process with intensity  $\lambda > 0$ . Then  $J_{N,n}(N)$  converges in m.s. and a.s. as  $n \rightarrow \infty$ . The limit of  $J_{N,n}(N)$ , denoted by  $\int_0^t N(s-) dN(s)$  or  $\int_0^t N_- dN$  and called the Itô integral of  $N_-$  with respect to  $N$  on  $[0, t]$ , is

$$\int_0^t N_-(s) dN(s) = \int_0^t N(s-) dN(s) = \frac{1}{2}(N(t)^2 - N(t)), \quad (4.9)$$

where  $N_-(s) = N(s-) = \lim_{u \uparrow s} N(u)$ . The Itô integral  $\int N_- dN$  coincides with the path by path Riemann-Stieltjes integral.  $\diamond$

*Proof* Let  $\alpha$  and  $\beta$  be real-valued functions such that  $\beta$  is a step function with jumps  $\beta_k$  at  $x_k$ ,  $k = 1, 2, \dots, n$ . If  $\alpha$  and  $\beta$  are not discontinuous from the right or from the left at each  $x_k$  simultaneously, then the Riemann-Stieltjes integral  $\int_a^b \alpha(x) d\beta(x)$  is given by  $\int_a^b \alpha(x) d\beta(x) = \sum_{k=1}^n \alpha(x_k) \beta_k$  ([1], Theorem 7.11). We have

$$\begin{aligned}
\int_0^t N(s-, \omega) dN(s, \omega) &= \sum_{k=1}^{N(t, \omega)} N(T_{k-1}(\omega), \omega) [N(T_k(\omega), \omega) - N(T_{k-1}(\omega), \omega)] \\
&= \sum_{k=1}^{N(t, \omega)} (k-1) = \frac{1}{2} (N(t, \omega)^2 - N(t, \omega)),
\end{aligned}$$

where  $T_k$ ,  $k = 1, 2, \dots$ , denote the jump times of  $N$  and  $T_0 = 0$ , that is, the path by path integral  $\int_0^t N(s-, \omega) dN(s, \omega)$  coincides with the Itô integral in agreement with a result in [6] (Theorem 17, p. 54), and  $J_{N,n}(N)$  converges a.s. to  $\int_0^t N_- dN$  as  $n \rightarrow \infty$ .

The m.s. convergence of  $J_{N,n}(N)$  to  $\int_0^t N_- dN = (N(t)^2 - N(t))/2$  follows by direct calculations and elementary arguments. We need to show that the expectation of the square of

$$J_{N,n}(N) - \frac{1}{2} (N(t)^2 - N(t)) = \frac{1}{2} \sum_k [\Delta N_k - (\Delta N_k)^2]$$

converges to zero as  $n \rightarrow \infty$ , where  $\Delta N_k = N(t_k) - N(t_{k-1})$ , that is, the sum

$$\begin{aligned}
&\sum_{k,l} E[(\Delta N_k - (\Delta N_k)^2)(\Delta N_l - (\Delta N_l)^2)] \\
&= \sum_{k,l} E[\Delta N_k \Delta N_l - \Delta N_k (\Delta N_l)^2 - \Delta N_l (\Delta N_k)^2 + (\Delta N_k)^2 (\Delta N_l)^2] \\
&= \lambda^4 \sum_{k,l} (\Delta t_k)^2 (\Delta t_l)^2 + 2\lambda^2 \sum_k (\Delta t_k)^2 + 4\lambda^3 \sum_k (\Delta t_k)^3
\end{aligned}$$

converges to zero as  $\Delta t_k = t_k - t_{k-1} \rightarrow 0$ . The first four moments of the increments of  $N$  in the above expression result from the cumulants  $\chi_r$  of  $\Delta N_k$  that are equal to  $\lambda \Delta t_k$  for any order  $r \geq 1$  and the relationships  $\mu_1 = \chi_1$ ,  $\mu_2 = \chi_2 + \chi_1^2$ ,  $\mu_3 = \chi_3 + 3\chi_1\chi_2 + \chi_1^3$ , and  $\mu_4 = \chi_4 + 3\chi_2^2 + 4\chi_1\chi_3 + 6\chi_1^2\chi_2 + \chi_1^4$  between cumulants and moments  $\mu_r = E[(\Delta N_k)^r]$  ([3], p. 377). The cumulant of order  $r$  of a random variable with characteristic function  $\varphi$  is  $i^{-r} d^r [\ln(\varphi(u))]/du^r$  for  $u = 0$ .  $\blacktriangle$

#### 4.4 Stochastic Integrals with Brownian Motion Integrators

We define stochastic integrals  $I(X) = \int_0^\tau X(s) dB(s)$  and  $I(X)(t) = \int_0^t X(s) dB(s)$ ,  $t \in [0, \tau]$ , representing random variables and stochastic processes, where  $X$  is a stochastic process satisfying some conditions and  $B$  denotes a standard Brownian motion.

Let  $X(t)$  and  $B(t)$ ,  $t \in [0, \tau]$ , be a real-valued stochastic process and a Brownian motion defined on a probability space  $(\Omega, \mathcal{F}, P)$ . It is assumed that  $X$  is  $\mathcal{F}_t = \sigma(B(s) : 0 \leq s \leq t)$ -adapted and  $\mathcal{B}[0, \tau] \times \mathcal{F}_\tau$ -measurable. Set

$$\begin{aligned}
\mathcal{H}_0^2 &= \mathcal{H}_0^2[0, \tau] = \{X : [0, \tau] \times \Omega \rightarrow \mathbb{R}, \text{ step process defined by (4.11)}\} \\
\mathcal{H}^2 &= \mathcal{H}^2[0, \tau] = \left\{ X : [0, \tau] \times \Omega \rightarrow \mathbb{R}, E \left[ \int_0^\tau X(t, \omega)^2 dt \right] < \infty \right\} \\
\mathcal{H} &= \mathcal{H}[0, \tau] = \left\{ X : [0, \tau] \times \Omega \rightarrow \mathbb{R}, P \left( \int_0^\tau X(t, \omega)^2 dt < \infty \right) = 1 \right\}
\end{aligned} \tag{4.10}$$

The members of  $\mathcal{H}_0^2$  have the form

$$X(t, \omega) = \sum_{i=1}^n A_i(\omega) 1(t_{i-1} < t \leq t_i), \tag{4.11}$$

where  $0 = t_0 < t_1 < \dots < t_n = \tau$  defines a partition of  $[0, \tau]$ ,  $A_i \in \mathcal{F}_{t_{i-1}}$ , and  $E[A_i^2] < \infty$ . The processes in  $\mathcal{H}_0^2$  have left continuous samples with right limits. We have  $\mathcal{H}_0^2 \subset \mathcal{H}^2 \subset \mathcal{H}$  since  $E \left[ \int_0^\tau X(t, \omega)^2 dt \right] = E \left[ \int_0^\tau \sum_{i,j=1}^n A_i A_j 1(t_{i-1} < t \leq t_i) 1(t_{j-1} < t \leq t_j) dt \right] = \sum_{i=1}^n E[A_i^2] \int_0^\tau 1(t_{i-1} < t \leq t_i) dt \leq \tau \sum_{i=1}^n E[A_i^2] < \infty$  for all  $X \in \mathcal{H}_0^2$  and  $P \left( \int_0^\tau X(t, \omega)^2 dt > a \right) \leq E \left[ \int_0^\tau X(t, \omega)^2 dt \right] / a \rightarrow 0$  as  $a \rightarrow \infty$  for all  $X \in \mathcal{H}^2$  by Chebyshev's inequality.

We define the Itô integral  $\int X dB$  for integrands  $X \in \mathcal{H}_0^2[0, \tau]$ , extend this definition to integrands  $X \in \mathcal{H}^2[0, \tau]$ , define an integral process  $I(X)(t)$ ,  $t \in [0, \tau]$ , for integrands  $X \in \mathcal{H}^2[0, \tau]$ , and extend resulting integrals to integrands  $X \in \mathcal{H}[0, \tau]$ . Our discussion is based on [4] (Chaps. 4, 5, and 6), [6] (Chap. 2), and [7] (Chaps. 6 and 7).

#### 4.4.1 Integrands in $\mathcal{H}_0^2$

The Itô integral for step process integrands can be introduced using elementary arguments. We define this integral and show that the mapping  $X \mapsto I(X)$  is an isometry.

**Definition 4.1** The stochastic integral for  $X \in \mathcal{H}_0^2$  is defined by

$$I(X) = \sum_{i=1}^n A_i (B(t_i) - B(t_{i-1})). \tag{4.12}$$

Note that  $\mathcal{H}_0^2$  is a linear space and that  $I(X)$  is random variable on  $(\Omega, \mathcal{F}, P)$  depending linearly on  $X$  (Exercise 4.4).

**Theorem 4.1** (Itô's isometry in  $\mathcal{H}_0^2$ ) *The equality*

$$\|I(X)\|_{L^2(dP)} = \|X\|_{L^2(dt \times dP)} \tag{4.13}$$

holds for all  $X \in \mathcal{H}_0^2$ , where  $\|\cdot\|_{L^2(dP)}$  and  $\|\cdot\|_{L^2(dt \times dP)}$  are the norms in  $L^2(dP) = L^2(\Omega, \mathcal{F}, P)$  and  $L^2(dt \times dP) = L^2([0, \tau] \times \Omega, \mathcal{B}[0, \tau] \times \mathcal{F}, \lambda \times P)$ .

*Proof* We have

$$\begin{aligned} \|X\|_{L^2(dt \times dP)}^2 &= E \left[ \int_0^\tau X(t, \omega)^2 dt \right] = \sum_{i=1}^n E[A_i^2] \int_0^\tau 1_{(t_{i-1} < t \leq t_i)} dt \\ &= \sum_{i=1}^n E[A_i^2] (t_i - t_{i-1}) \end{aligned}$$

and

$$\begin{aligned} \|I(X)\|_{L^2(dP)}^2 &= \sum_{i,j=1}^n E[A_i A_j \Delta B_i \Delta B_j] = \sum_{i=1}^n E[A_i^2 \Delta B_i^2] \\ &= \sum_{i=1}^n E\{E[A_i^2 \Delta B_i^2 \mid \mathcal{F}_{t_{i-1}}]\} = \sum_{i=1}^n E\{A_i^2 E[\Delta B_i^2 \mid \mathcal{F}_{t_{i-1}}]\} = \sum_{i=1}^n E[A_i^2] (t_i - t_{i-1}) \end{aligned}$$

by the definition of  $I(X)$ , properties of  $\{A_i\}$  and  $B$ , and  $E[A_i A_j \Delta B_i \Delta B_j] = 0$  for  $i \neq j$ , where  $\Delta B_i = B(t_i) - B(t_{i-1})$ . For example, if  $i > j$ ,  $E[A_i A_j \Delta B_i \Delta B_j] = E\{E[A_i A_j \Delta B_i \Delta B_j \mid \mathcal{F}_{t_{i-1}}]\} = E\{A_i A_j \Delta B_j E[\Delta B_i \mid \mathcal{F}_{t_{i-1}}]\}$  and  $E[\Delta B_i \mid \mathcal{F}_{t_{i-1}}] = 0$  by the martingale property of Brownian motion.  $\blacktriangle$

Itô's isometry shows that  $I : \mathcal{H}_0^2 \rightarrow L^2(dP)$  maps  $\mathcal{H}_0^2$  continuously in  $L^2(dP)$ . Since  $I$  preserves distance, Cauchy sequences in  $\mathcal{H}_0^2$  are mapped into Cauchy sequences in  $L^2(dP)$ .

#### 4.4.2 Integrands in $\mathcal{H}^2$

First, the Itô integral  $I(X) = \int_0^\tau X(t) dB(t)$  is constructed for  $X \in \mathcal{H}^2$  by using the fact that  $\mathcal{H}_0^2$  is dense in  $\mathcal{H}^2$  ([7], Lemma 6.2). Hence, for any  $X \in \mathcal{H}^2$  there exists a sequence of step processes  $X_n \in \mathcal{H}_0^2$  such that  $\|X_n - X\|_{L^2(dt \times dP)} \rightarrow 0$  as  $n \rightarrow \infty$ . Second, the integral  $I(X)$  is extended to a stochastic process  $I(X)(t)$ ,  $t \in [0, \tau]$ .

**Theorem 4.2** *Let  $X_n \in \mathcal{H}_0^2$ ,  $n = 1, 2, \dots$ , be a sequence of step processes converging to  $X \in \mathcal{H}^2$ . The stochastic integral  $I(X) = \int_0^\tau X(t) dB(t)$  is the m.s. limit*

$$I(X) = \lim_{n \rightarrow \infty} I(X_n), \quad (4.14)$$

where  $I(X_n) = \int_0^\tau X_n(t) dB(t)$ .

*Proof* Since  $\|X_n - X\|_{L^2(dt \times dP)} \rightarrow 0$  as  $n \rightarrow \infty$ , the sequence of step processes is Cauchy in  $L^2(dt \times dP)$ . The Itô isometry implies that the sequence of stochastic

integrals  $\{I(X_n)\}$  is Cauchy in  $L^2(dP)$  so that it is convergent with limit in  $L^2(dP)$  since this space is complete. Consider another sequence  $\{X'_n\}$  of step processes in  $\mathcal{H}_0^2$  that converges to  $X$  in  $L^2(dt \times dP)$ . We have  $\|X_n - X'_n\|_{L^2(dt \times dP)} \rightarrow 0$  by the triangle inequality, which implies  $\|I(X_n) - I(X'_n)\|_{L^2(dP)} \rightarrow 0$  as  $n \rightarrow \infty$  by Itô's isometry. Hence, the definition of  $I(X)$  in (4.14) is meaningful since it is independent of the particular sequence of step processes  $\{X_n\}$  used for its definition.  $\blacktriangle$

**Theorem 4.3** *For  $X \in \mathcal{H}^2$  arbitrary, we have*

$$\|I(X)\|_{L^2(dP)} = \|X\|_{L^2(dt \times dP)} \text{ and} \quad (4.15)$$

$$E\left[\left(\int_s^t X(u) dB(u)\right)^2 \mid \mathcal{F}_s\right] = E\left[\int_s^t X(u)^2 du \mid \mathcal{F}_s\right], \quad (4.16)$$

where  $0 \leq s \leq t \leq \tau$ .

*Proof* Let  $X_n \in \mathcal{H}_0^2$ ,  $n = 1, 2, \dots$ , be a sequence of step processes converging to  $X \in \mathcal{H}^2$ . We have  $\|X_n\| \leq \|X_n - X\| + \|X\|$  and  $\|X\| \leq \|X - X_n\| + \|X_n\|$  by the triangle inequality implying  $|\|X_n\| - \|X\|| \leq \|X - X_n\|$ , where all norms are in  $L^2(dt \times dP)$ . Hence,  $\|X_n\| \rightarrow \|X\|$  as  $n \rightarrow \infty$  since  $\{X_n\}$  converges to  $X$  in  $L^2(dt \times dP)$ . Similar arguments give the convergence  $\|I(X_n)\| \rightarrow \|I(X)\|$ ,  $n \rightarrow \infty$ , where the norms are in  $L^2(dP)$ . These properties and the Itô isometry in (4.13) give (4.15).

For (4.16) it is sufficient to prove  $E[1_A (\int_s^t X(u) dB(u))^2] = E[1_A \int_s^t X(u)^2 du]$  for all  $A \in \mathcal{F}_s$ , that is, (4.15) with  $\tilde{X} = 1_A X$  in place of  $X$ . Since  $\tilde{X} \in \mathcal{H}^2$ , we have  $\|I(\tilde{X})\|_{L^2(dP)} = \|\tilde{X}\|_{L^2(dt \times dP)}$ , that is, (4.16).  $\blacktriangle$

We now construct the integral  $I(X)(t) = \int_0^t X(s) dB(s)$ ,  $t \in [0, \tau]$ , with integrands  $X \in \mathcal{H}^2[0, \tau]$ . The construction must account for the fact that the Itô integral  $I(X)$  in (4.14) has been defined as a m.s. limit of a sequence of integrals with step process integrands, so that it takes arbitrary values on a set of zero probability measure. For a fixed  $t \in [0, \tau]$ , the definition

$$\begin{aligned} I(X)(t) &= \int_0^t X(s) dB(s) = \int_0^\tau 1(0 \leq s \leq t) X(s) dB(s) \\ &= \int_0^\tau \tilde{X}(s) dB(s), \quad t \in [0, \tau], \end{aligned} \quad (4.17)$$

is valid since  $\tilde{X}(s) = 1(0 \leq s \leq t)X(s)$  is in  $\mathcal{H}^2[0, \tau]$ , so that  $I(X)(t)$  is ambiguous on a set  $A_t \in \mathcal{F}$  with  $P(A_t) = 0$ . This definition can be extended to a countable number of times  $t_i \in [0, \tau]$  since the sets  $A_{t_i}$  on which the integrals  $I(X)(t_i)$  take arbitrary values are measurable and have probability 0, so that the integral at all these times is not uniquely defined on  $\cup_i A_{t_i}$  and  $P(\cup_i A_{t_i}) \leq \sum_i P(A_{t_i}) = 0$ . However, (4.17) cannot be extended to all  $t$  in  $[0, \tau]$  since  $[0, \tau]$  is uncountable, so that  $\cup_{t \in [0, \tau]} A_t$

may not be an event and, even if it is, its measure may be strictly positive. This difficulty is resolved by the following theorem.

**Theorem 4.4** *For any  $X \in \mathcal{H}^2[0, \tau]$ , there exists a continuous  $\mathcal{F}_t$ -martingale  $M(t)$ ,  $t \in [0, \tau]$ , such that*

$$P\left(\left\{\omega \in \Omega : M(t, \omega) = \int_0^\tau 1(0 \leq s \leq t)X(s, \omega) dB(s, \omega)\right\}\right) = 1 \quad (4.18)$$

for each  $t \in [0, \tau]$ , that is, the process  $\int_0^\tau 1(0 \leq s \leq t)X(s) dB(s)$  has a modification that is a continuous martingale with respect to the Brownian filtration ([7], Theorem 6.2).

The theorem shows that the process  $I(X)(t)$  in (4.17) defined by an Itô's integral is useful provided we work with a modification of  $I(X)(t)$  that is a continuous martingale with respect to the Brownian filtration.

*Example 4.5* The process

$$M(t) = \left(\int_0^t X(u) dB(u)\right)^2 - \int_0^t X(u)^2 du, \quad t \geq 0, \quad (4.19)$$

is a martingale.  $M(t)$  has finite mean, is  $\mathcal{F}_t$ -adapted as a function of  $B(s)$ ,  $0 \leq s \leq t$ , and, for  $t > s$ , we have

$$\begin{aligned} E[M(t) \mid \mathcal{F}_s] &= M(s) + E\left[\left(\int_s^t X(u) dB(u)\right)^2 - \int_s^t X(u)^2 du \right. \\ &\quad \left. + 2\left(\int_0^s X(u) dB(u)\right)\left(\int_s^t X(u) dB(u)\right) \mid \mathcal{F}_s\right] \\ &= M(s) + E\left[\left(\int_s^t X(u) dB(u)\right)^2 - \int_s^t X(u)^2 du \mid \mathcal{F}_s\right]. \end{aligned}$$

The latter equality holds since  $\int_0^s X(u) dB(u) \in \mathcal{F}_s$ ,  $\int_0^t X(u) dB(u)$  is a martingale by Theorem 4.4 so that  $E\left[\left(\int_s^t X(u) dB(u)\right) \mid \mathcal{F}_s\right] = 0$ . Since  $E\left[\left(\int_s^t X(u) dB(u)\right)^2 - \int_s^t X(u)^2 du \mid \mathcal{F}_s\right] = 0$  by Theorem 4.3, we have  $E[M(t) \mid \mathcal{F}_s] = M(s)$ .  $\diamond$

*Example 4.6* The special case of  $M(t)$  in (4.19) with  $X(u) = 1$  shows that the process  $M(t) = \left(\int_0^t dB(u)\right)^2 - \int_0^t du = B(t)^2 - t$  is a martingale. We can also see that  $M(t) = B(t)^2 - t$  is martingale by direct calculations. By its definition,  $M(t)$  has finite mean and is  $\mathcal{F}_t$ -adapted. For  $t > s$ ,  $E[M(t) \mid \mathcal{F}_s] = E[B(t)^2 \mid \mathcal{F}_s] - t = E[(B(t) - B(s))^2 + B(s)^2 + 2(B(t) - B(s))B(s) \mid \mathcal{F}_s] - t = E[(B(t) - B(s))^2] + B(s)^2 + 2B(s)E[B(t) - B(s)] - t = (t - s) + B(s)^2 - t = M(s)$ , since  $B(t) - B(s)$  is independent of  $\mathcal{F}_s$  and  $B(s) \in \mathcal{F}_s$ .  $\diamond$

### 4.4.3 Integrands in $\mathcal{H}$

The set of integrands  $\mathcal{H}^2$  is too restrictive. For example, the continuous function  $X(t) = \exp(B(t)^4)$  of a Brownian motion is not in  $\mathcal{H}^2$  but is in  $\mathcal{H}$ . Localization sequences are used to extend the Itô integral defined in the previous section to integrands in  $\mathcal{H}$ .

**Definition 4.2** An increasing sequence of stopping times  $T_n, n = 1, 2, \dots$ , is an  $\mathcal{H}^2[0, \tau]$  localizing sequence for  $X \in \mathcal{H}[0, \tau]$  if  $X_n(t) = X(t) 1(t \leq T_n) \in \mathcal{H}^2[0, \tau]$  for all  $n$  and  $P(\bigcup_{n=1}^{\infty} \{T_n = \tau\}) = 1$ . Recall that a random variable  $T : \Omega \rightarrow [0, \infty]$  is an  $\mathcal{F}_t$ -stopping time if  $\{T \leq t\} \in \mathcal{F}_t$  for all  $t \geq 0$ .

The random times

$$T_n(\omega) = \inf \left\{ t : \int_0^t X(s, \omega)^2 ds \geq n \text{ or } t \geq \tau \right\}, \quad n = 1, 2, \dots, \quad (4.20)$$

define a localizing sequence for  $X \in \mathcal{H}$  since  $P(\{\omega : \int_0^\tau X(s, \omega)^2 ds < \infty\}) = 1$  and  $\bigcup_{n=1}^{\infty} \{\omega : T_n(\omega) = \tau\} = \{\omega : \int_0^\tau X(s, \omega)^2 ds < \infty\}$ . Note also that the processes  $X_n(t) = X(t) 1(0 \leq t \leq T_n)$  are in  $\mathcal{H}^2[0, \tau]$  and have the property  $\|X_n\|_{L^2(dt \times dP)}^2 \leq n$  by construction.

We define the stochastic process  $\int_0^t X(s) dB(s), X \in \mathcal{H}$ , in two steps. First, let  $M_n(t), t \in [0, \tau]$ , be the continuous  $\mathcal{F}_t$ -martingale in Theorem 4.4, which constitutes a modification of  $\int_0^\tau X_n(s) dB(s)$ . Second, define  $\int_0^\tau 1(0 \leq s \leq t) X(s) dB(s)$  as the limit of  $M_n(t)$  as  $n \rightarrow \infty$ , that is,

$$P\left(\int_0^\tau 1(0 \leq s \leq t) X(s) dB(s) = \lim_{n \rightarrow \infty} M_n(t)\right) = 1 \quad \text{for all } t \in [0, \tau]. \quad (4.21)$$

The definition in (4.21) is justified since  $M_m(t) = M_n(t), n \geq m$ , for almost all  $\omega \in \{t \leq T_m\}$  ([7], Proposition 7.2). Hence, there is a continuous process  $M(t), 0 \leq t \leq \tau$ , such that  $P(M(t) = \lim_{n \rightarrow \infty} M_n(t)) = 1$  for all  $t \in [0, \tau]$  ([7], Proposition 7.3). If  $M'_n(t)$  corresponds to another localizing sequence  $\{T'_n\}$  for  $X \in \mathcal{H}$ , then  $\lim_{n \rightarrow \infty} M_n(t) = \lim_{n \rightarrow \infty} M'_n(t)$  a.s. at each  $t \in [0, \tau]$  ([7], Proposition 7.4). Also, if  $T$  is a stopping time and  $X, Y \in \mathcal{H}$  such that  $X(s, \omega) = Y(s, \omega)$  for all  $s \in [0, T]$  then  $\int_0^\tau 1(0 \leq s \leq t) X(s) dB(s) = \int_0^\tau 1(0 \leq s \leq t) Y(s) dB(s)$  for almost all  $\omega \in \{t \leq T\}$  ([7], Proposition 7.5).

**Example 4.7** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a continuous function and consider the Itô integral  $I(X) = \int_0^\tau X(t) dB(t)$  with integrand  $X(t) = f(B(t)) \in \mathcal{H}$ . Then

$$\int_0^\tau f(B(t)) dB(t) = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(B(t_{i-1})) (B(t_i) - B(t_{i-1})) \quad (4.22)$$

in probability, where  $t_i = i\tau/n, i = 0, 1, \dots, n$  ([7], Theorem 7.1). This extends results in Example 4.1 to functions  $f$  other than the identity function.  $\diamond$



*Proof* For  $a > 0$ , note that  $T_a = \inf\{t : |B(t)| > a \text{ or } t \geq \tau\}$  is a localizing sequence for  $f(B(t))$ ,  $f_a(B(t)) = f(B(t)) 1(|B(t)| \leq a)$  is in  $\mathcal{H}_0^2[0, \tau]$ , and there exists a sequence  $f_{a,n}(B(t))$  in  $\mathcal{H}_0^2[0, \tau]$  such that  $\|f_{a,n} - f_a\|_{L^2(dt \times dP)} \rightarrow 0$  as  $n \rightarrow \infty$ . That  $f_{a,n}(B(t)) = \sum_{i=1}^n f_a(B(t_{i-1})) 1(t_{i-1} < t \leq t_i)$  is an approximating sequence for  $f_a(B(t))$  follows from

$$\begin{aligned} & E \left[ \int_0^\tau (f_{a,n}(B(t)) - f_a(B(t)))^2 dt \right] \\ &= E \left[ \int_0^\tau \sum_{i=1}^n (f_a(B(t_{i-1})) - f_a(B(t)))^2 1(t_{i-1} < t \leq t_i) dt \right] \\ &\leq \frac{\tau}{n} \sum_{i=1}^n E \left[ \sup_{t \in (t_{i-1}, t_i]} (f_a(B(t_{i-1})) - f_a(B(t)))^2 \right] \end{aligned}$$

since the random variables  $\sup_{t \in (t_{i-1}, t_i]} (f_a(B(t_{i-1})) - f_a(B(t)))^2$  are bounded and  $E \left[ \sup_{t \in (t_{i-1}, t_i]} (f_a(B(t_{i-1})) - f_a(B(t)))^2 \right] \rightarrow 0$  as  $n \rightarrow \infty$  a.s. by the uniform continuity of  $B$  on  $[0, \tau]$  and dominated convergence.

The Itô isometry implies  $\int_0^t f_a(B(s)) dB(s) = \lim_{n \rightarrow \infty} \sum_{i=1}^n f_{a,n}(B(t_{i-1})) (B(t_i) - B(t_{i-1}))$ , where the convergence is in  $L^2(dP)$ . It remains to show that the probability of event  $A_n(\varepsilon) = \{|\sum_{i=1}^n f(B(t_{i-1})) (B(t_i) - B(t_{i-1})) - \int_0^\tau f(B(s)) dB(s)| > \varepsilon\}$  converges to 0 as  $n \rightarrow \infty$ . We have  $P(A_n(\varepsilon)) = P(A_n(\varepsilon) \cap \{T_a < \tau\}) + P(A_n(\varepsilon) \cap \{T_a = \tau\}) \leq P(T_a < \tau) + P(A_n(\varepsilon) \cap \{T_a = \tau\})$ ,  $P(T_a < \tau) \rightarrow 0$  as  $a \rightarrow \infty$ , and the probability of  $A_n(\varepsilon) \cap \{T_a = \tau\}$  can be bounded by  $\|\sum_{i=1}^n f_a(B(t_{i-1})) (B(t_i) - B(t_{i-1})) - \int_0^\tau f_a(B(s)) dB(s)\|_{L^2(dP)}^2 / \varepsilon^2$ , which converges to 0 as  $n \rightarrow \infty$ .  $\blacktriangle$

**Example 4.8** The output of linear filters to Gaussian white noise is a special type of Itô's integral with deterministic integrand. The solution  $M(t)$  of a linear filter at a time  $t > 0$  defined by  $dM(t) = -\alpha M(t) dt + dB(t)$ ,  $\alpha > 0$ , is  $M(t) = \int_0^t \exp(-\alpha(t-s)) dB(s)$  for  $M(0) = 0$ , that is, an Itô integral with deterministic integrand. The process  $M(t)$  has mean 0 and variance  $E[M(t)^2] = \int_0^t \exp(-2\alpha(t-s)) ds = (1 - \exp(-2\alpha t)) / (2\alpha)$ .  $\diamond$

## 4.5 Stochastic Integrals with Martingale Integrators

We consider stochastic integrals  $\int_0^\tau X(t) dM(t)$  with integrands  $X$  satisfying some conditions and integrators  $M$  that are square integrable martingale with right continuous samples that have left limits. An example is the integral  $\int \tilde{C}_- d\tilde{C}$ , where  $\tilde{C}(t) = \sum_{k=1}^\infty Y_k 1(t \geq T_k) - \lambda t E[Y_1]$ ,  $t \geq 0$ , is the compensated compound Poisson process,  $\{T_k\}$  are the jump times of a homogeneous Poisson process  $N(t)$  with intensity  $\lambda > 0$ ,  $\{Y_k\}$  are iid random variables with finite variance, and  $\tilde{C}_-(t) = \lim_{s \uparrow t} \tilde{C}(s)$ .

Example 4.4 shows that Riemann-Stieltjes integrals can be defined for right continuous integrators and left continuous integrands provided their discontinuities are not coincident. Since martingales admit right continuous modifications and  $M$  in  $\int_0^\tau X(t) dM(t)$  is assumed to be a right continuous martingale, it is natural to require that the integrand  $X$  has left continuous samples.

Let  $(\Omega, \mathcal{F}, P)$  be a probability space with a filtration  $\mathcal{F}_t$ ,  $t \geq 0$ . Denote by  $\mathcal{L}$  and  $\mathcal{D}$  the collection of all jointly measurable,  $\mathcal{F}_t$ -adapted real-valued stochastic processes that have left continuous samples with right limits and right continuous samples with left limits, respectively.

**Definition 4.3** Let  $\mathcal{P}$  and  $\mathcal{O}$  be the smallest  $\sigma$ -fields on  $[0, \tau] \times \Omega$  with respect to which all processes  $X(t)$ ,  $t \in [0, \tau]$ , in  $\mathcal{L}$  and  $\mathcal{D}$  are jointly measurable. If the mapping  $(t, \omega) \mapsto X(t, \omega)$  is  $\mathcal{P}$ -measurable,  $X$  is said to be predictable. If this mapping is  $\mathcal{O}$ -measurable,  $X$  is said to be optional.

Note that (1) the  $\sigma$ -field  $\mathcal{P}$  is generated by subsets of  $[0, \tau] \times \Omega$  of the type  $(s, t] \times A$ ,  $0 \leq s < t \leq \tau$ ,  $A \in \mathcal{F}_s$ , and  $\{0\} \times A$ ,  $A \in \mathcal{F}_0$  ([4], Sect. 6.3), (2) predictable and optional processes are measurable since the  $\sigma$ -fields  $\mathcal{P}$  and  $\mathcal{O}$  are included in  $\mathcal{B}[0, \tau] \times \mathcal{F}$ , (3) predictable processes allow “a peek into the future” since their samples are left continuous, and (4) any predictable process is also optional, that is,  $\mathcal{P} \subseteq \mathcal{O}$ . Some of these concepts have been already discussed (Definition 3.35).

Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\mathcal{F}_t$ ,  $0 \leq t \leq \tau$ , a filtration on this space with respect to which the integrator  $M$  of  $\int_0^\tau X dM$  is a martingale with right continuous samples and  $X$  is a predictable process. It is assumed that the filtration  $\mathcal{F}_t$  satisfies the usual hypothesis, that is,  $\mathcal{F}_0$  contains all sets of  $P$ -measure 0 (completeness) and  $\mathcal{F}_t = \mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$  for all  $t$  (right continuity).

We have seen that discrete time submartingales admit the Doob decomposition (Theorem 2.18). A similar decomposition, referred to as the Doob-Meyer decomposition, is available for continuous time submartingales. We state a simpler version of this decomposition that is needed for the construction of stochastic integrals with martingale integrators. Technical details on the Doob-Meyer decomposition can be found in [6] (Sect. 3.2) and [2] (Theorem 5.1).

**Theorem 4.5** (Doob-Meyer decomposition). *If  $M$  is a square integrable martingale with samples in  $\mathcal{D}$ , there exists a unique decomposition*

$$M(t)^2 = L(t) + \langle M \rangle(t), \quad 0 \leq t \leq \tau, \quad (4.23)$$

where  $L(t) \in \mathcal{D}$  is a martingale and  $\langle M \rangle(t)$ , referred to as the compensator of  $M(t)^2$ , is a predictable process with increasing samples such that  $\langle M \rangle(0) = 0$  and  $E[\langle M \rangle(t)] < \infty$  for all  $t \in [0, \tau]$  ([2], Theorem 5.1).

The compensator  $\langle M \rangle(t)$  defined by the Doob-Meyer decomposition in (4.23) is unique. If  $M$  is continuous, the processes  $\langle M \rangle(t)$  and  $[M, M](t)$  are indistinguishable ([2], Sect. 2.6), where  $[M, M] = [M]$  denotes the quadratic variation of  $M$  that is defined in the following section. If  $M$  is a Brownian motion  $B$ , then  $[M, M](t) = \langle M \rangle(t) = t$ , where the last equality holds since  $[B](t) = t$  (Sect. 3.7.6.1).

*Example 4.9* Let  $M(t) = N(t) - \lambda t$ ,  $t \geq 0$ , where  $N$  is a Poisson process with intensity  $\lambda > 0$ . The compensator of  $M(t)$  is  $\langle M \rangle(t) = \lambda t$ .  $\diamond$

*Proof* Since  $M(t)$  is a square integrable martingale with samples in  $\mathcal{D}$ , we need to show that  $L(t) = M(t)^2 - \lambda t$  is a martingale with respect to the filtration  $\mathcal{F}_t$  generated by  $N$  that has samples in  $\mathcal{D}$ . We have  $E[|L(t)|] < \infty$  and  $L(t) \in \mathcal{F}_t$ ,  $t \geq 0$ , by construction, and

$$\begin{aligned} E[M(t)^2 \mid \mathcal{F}_s] &= E \left[ (N(t) - N(s) - \lambda(t-s) + (N(s) - \lambda s))^2 \mid \mathcal{F}_s \right] \\ &= E \left[ (N(t) - N(s) - \lambda(t-s))^2 \mid \mathcal{F}_s \right] \\ &\quad + 2E \left[ (N(t) - N(s) - \lambda(t-s))(N(s) - \lambda s) \mid \mathcal{F}_s \right] \\ &\quad + E \left[ (N(s) - \lambda s)^2 \mid \mathcal{F}_s \right] = \lambda(t-s) + (N(s) - \lambda s)^2 = M(s)^2 + \lambda(t-s). \end{aligned}$$

for  $s \leq t$ . Hence,  $M^2(t) - \lambda t$ ,  $t \geq 0$ , is a martingale with samples in  $\mathcal{D}$ . Since the Doob-Meyer decomposition is unique, we have  $\langle M \rangle(t) = \lambda t$ .  $\blacktriangle$

Let  $\mathcal{H}_{\text{pred}}^2[0, \tau] = \mathcal{H}_{\text{pred}}^2$  be the collection of predictable processes  $X(t)$ ,  $0 \leq t \leq \tau$ , with the property  $E \left[ \int_0^\tau |X(t)|^2 d\langle M \rangle(t) \right] < \infty$ . For the special case  $M = B$ , we have  $d\langle M \rangle(t) = dt$  and  $\mathcal{H}_{\text{pred}}^2 = \mathcal{H}^2$ . The construction of the Itô integral  $\int_0^\tau X(t) dM(t)$  with martingale integrator is similar to that with Brownian motion integrator. First,  $\int_0^\tau X(t) dM(t)$  is defined for integrands  $X \in \mathcal{H}_{\text{pred}}^2$  that are step processes. Second, the resulting integral is extended to arbitrary  $X \in \mathcal{H}_{\text{pred}}^2$ . Third, the Itô integral is defined for integrands  $X$  in the class  $\mathcal{H}_{\text{pred}}$  of predictable processes with the property  $\int_0^\tau |X(t)|^2 d\langle M \rangle(t) < \infty$  a.s., rather than  $E \left[ \int_0^\tau |X(t)|^2 d\langle M \rangle(t) \right] < \infty$ . Note that  $\mathcal{H}_{\text{pred}}$  is a larger class of integrands, that is,  $\mathcal{H}_{\text{pred}} \supset \mathcal{H}_{\text{pred}}^2$ .

For a step process  $X$  in  $\mathcal{H}_{\text{pred}}^2$  of the type in (4.11), the Itô integral is

$$I(X) = \sum_{i=1}^n A_i (M(t_i) - M(t_{i-1})), \quad (4.24)$$

that is, (4.12) with  $M$  in place of  $B$ . The following result resembles Theorem 4.1.

**Theorem 4.6** For  $I(X)$  in (4.24) we have

$$E[|I(X)|^2] = E \left[ \int_0^\tau |X(t)|^2 d\langle M \rangle(t) \right]. \quad (4.25)$$

*Proof* Set  $\Delta M_i = M(t_i) - M(t_{i-1})$  and note that

$$\begin{aligned} E[|I(X)|^2] &= \sum_{i=1}^n E \left[ A_i^2 (\Delta M_i)^2 \right] + \sum_{i,j=1, i \neq j}^n E \left[ A_i A_j \Delta M_i \Delta M_j \right] \\ &= \sum_{i=1}^n E \left[ A_i^2 (\Delta M_i)^2 \right] \end{aligned}$$

since  $E[A_i A_j \Delta M_i \Delta M_j] = E\{A_i A_j \Delta M_j E[\Delta M_i \mid \mathcal{F}_{t_{i-1}}]\}$  for  $i > j$  and the conditional expectation  $E[\Delta M_i \mid \mathcal{F}_{t_{i-1}}]$  is zero. This observation implies  $E[|I(X)|^2] = \sum_{i=1}^n E\{A_i^2 E[\langle M \rangle(t_i) - \langle M \rangle(t_{i-1}) \mid \mathcal{F}_{t_{i-1}}]\} = E\left[\int_0^\tau |X(t)|^2 d\langle M \rangle(t)\right]$  since  $M(t)^2 = L(t) + \langle M \rangle(t)$  by the Doob-Meyer decomposition in (4.23) and  $L(t)$  is an  $\mathcal{F}_t$ -martingale so that  $E[L(t_i) - L(t_{i-1}) \mid \mathcal{F}_{t_{i-1}}] = 0$ .  $\blacktriangle$

Let  $X \in \mathcal{H}_{\text{pred}}^2$  and let  $\{X_n\}$  be a sequence of step processes in  $\mathcal{H}_{\text{pred}}^2$  such that  $E\left[\int_0^\tau |X(t) - X_n(t)|^2 d\langle M \rangle(t)\right] \rightarrow 0$  as  $n \rightarrow \infty$ . Theorem 4.6 shows that  $\{I(X_n)\}$  is a Cauchy sequence in  $L^2(dP)$  so that it has a unique limit in this space denoted by  $I(X) = \lim_{n \rightarrow \infty} I(X_n)$ . Since  $I(X)$  does not depend on the particular sequence  $\{X_n\}$ , it is well defined. The limit  $I(X) = \lim_{n \rightarrow \infty} I(X_n)$  is called the stochastic integral of  $X$  with respect to martingale  $M$ .

*Example 4.10* If  $M(t)$  is a continuous square integrable martingale, then the stochastic integral with  $M$  as integrand and integrator is  $\int_0^t M(s) dM(s) = (M(t)^2 - M(0)^2 - \langle M \rangle(t))/2 = (M(t)^2 - M(0)^2 - [M](t))/2$ .  $\diamond$

*Proof* Set  $J_{M,n}(M) = \sum_{i=1}^n M(t_{i-1}) \Delta M_i$  as in Example 4.1, where  $\Delta M_i = M(t_i) - M(t_{i-1})$  and  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 < t_1 < \dots < t_n = t$ , is a sequence of partitions of  $[0, t]$  with intermediate points  $t'_i = t_{i-1}$  such that  $\Delta(p_n) = \max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$  as  $n \rightarrow \infty$ . Our objective is to show the convergence  $J_{M,n}(M) - M(0)^2/2 \rightarrow (M(t)^2 - M(0)^2 - \langle M \rangle(t))/2$ . Since  $J_{M,n}(M) = M(t)^2/2 - \sum_{k=1}^n (\Delta M_i)^2/2$ , it is sufficient to show the convergence  $\sum_{i=1}^n (\Delta M_i)^2 \rightarrow [M](t) = \langle M \rangle(t)$ , where the latter equality is valid since  $M$  is assumed to have continuous samples. This convergence is in probability and follows from Theorem 4.11 stated in the following section.  $\blacktriangle$

We conclude this section with a theorem given without proof that is useful for calculations, and a brief outline on the construction of the Itô integral for integrands  $X \in \mathcal{H}_{\text{pred}}$ .

**Theorem 4.7** If  $X \in \mathcal{H}_{\text{pred}}^2$ , the stochastic process  $I(X)(t) = \int_0^t X(s) dM(s)$ ,  $t \in [0, \tau]$ , is a martingale and  $E[I(X)(t)^2] = E\left[\int_0^t |X(s)|^2 d\langle M \rangle(s)\right]$ . Moreover, the process  $I(X)(t)$  has right continuous samples with left limits and the compensator of  $(I(X)(t))^2$  is  $\langle I(X) \rangle(t) = \int_0^t |X(s)|^2 d\langle M \rangle(s)$  ([4], Theorems 6.5.8 and 6.5.9).

Let  $X \in \mathcal{H}_{\text{pred}}$  and set  $X_n(t) = X(t)1\left(\int_0^t |X(s)|^2 d\langle M \rangle(s) \leq n\right)$ . Since  $X_n \in \mathcal{H}_{\text{pred}}^2$ , the Itô integral  $I(X_n)$  is defined. It can be shown that the sequence  $\{I(X_n)\}$  is convergent in probability and its limit  $I(X) = \lim_{n \rightarrow \infty} I(X_n)$  is well defined ([4], Chap. 6). The limit  $I(X)$  is called the Itô integral of  $X \in \mathcal{H}_{\text{pred}}$  with respect to a martingale  $M$ . The following result is for a special type of stochastic integral  $I(X)$ .

**Theorem 4.8** If  $X \in \mathcal{H}_{\text{pred}}$ ,  $M$  is a square integrable martingale, and both  $X$  and  $M$  have continuous samples, then  $\int_0^\tau X(t) dM(t) = \lim_{\Delta(p_n) \rightarrow 0} \sum_{i=1}^n X(t_{i-1}) (M(t_i) - M(t_{i-1}))$  in probability, where  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 < t_1 < \dots < t_n = \tau$ , is

a sequence of partitions of  $[0, \tau]$  and  $\Delta(p_n) = \max_{1 \leq i \leq n} (t_i - t_{i-1})$  ([4], Theorem 6.6.5).

## 4.6 Stochastic Integrals with Semimartingale Integrators

Stochastic integrals with Brownian motions and martingales integrators considered in previous sections are extended to integrals with semimartingale integrators.

**Definition 4.4** A process  $Y \in \mathcal{D}$  is a semimartingale if and only if it admits the representation

$$Y(t) = Y(0) + M(t) + A(t), \quad (4.26)$$

where  $M(0) = A(0) = 0$ ,  $M(t)$  is a local martingale, and  $A(t)$  is a process in  $\mathcal{D}$  of finite variation on compacts ([6], Theorem 14, p. 105 and Theorem 22, p. 114).

Recall that  $M$  is a local martingale if there exists an increasing sequence  $T_n$ ,  $n = 1, 2, \dots$ , of stopping times such that  $\lim_{n \rightarrow \infty} T_n = +\infty$  a.s. and  $M(t \wedge T_n)$  is a martingale for each  $n$  (Definition 3.34). A process is said to be of finite variation on compacts if almost all its samples are of finite variation on each compact of  $\mathbb{R}$ .

*Example 4.11* The square of a Brownian motion  $B(t)$  is a semimartingale since it admits the representation  $Y(t) = B(t)^2 = Y(0) + A(t) + M(t)$  with  $Y(0) = 0$ ,  $A(t) = t$ , and  $M(t) = 2 \int_0^t B(s) dB(s)$ .  $\diamond$

*Proof* The representation of  $Y(t)$  results from (4.6). That  $M(t)$  is a martingale follows from the construction of Itô's integral with Brownian integrator. Note also that  $A(t) = t$  is of finite variation on compacts and continuous, so that  $A \in \mathcal{D}$ , and that  $A(0) = 0$ . Also,  $M(0) = 0$  by its definition.  $\blacktriangle$

*Example 4.12* Compound Poisson processes with integrable jumps are semimartingales.  $\diamond$

*Proof* Let  $C(t) = \sum_{k=1}^{N(t)} Y_k$ ,  $t \geq 0$ , be a compound Poisson process, where  $N(t)$  is a Poisson process with intensity  $\lambda > 0$  and  $\{Y_k\}$  denote iid random variables with finite mean. The process  $C(t)$  is a semimartingale since it admits the representation  $C(t) = M(t) + A(t)$ , where  $M(t) = C(t) - \lambda E[Y_1]t$  is a martingale and  $A(t) = \lambda E[Y_1]t$  is adapted and of finite variation.  $\blacktriangle$

*Example 4.13* The approximate representation  $\tilde{L}_{\alpha,a}(t) = \sigma(a)B(t) + C_{\alpha,a}(t)$  of an  $\alpha$ -stable process  $L_\alpha(t)$  in (3.62) is a semimartingale for  $\alpha \in (1, 2]$  since  $B(t)$  is a martingale and  $C_{\alpha,a}(t)$  is a semimartingale.  $\diamond$

We now consider stochastic integrals  $\int X dY$ , where the integrand  $X$  is a process in  $\mathcal{L}$  and the integrator  $Y$  is a semimartingale. The construction of these integrals is similar to that of Itô's integrals with Brownian motion and martingale integrators. Stochastic integrals  $\int X dY$  are first defined for integrands  $X \in \mathcal{L}$  that are step

processes. This definition is subsequently extended to arbitrary integrands  $X \in \mathcal{L}$ . A comprehensive discussion on stochastic integrals with semimartingale integrators can be found in [6] (Sect. 2.4).

**Example 4.14** Consider the stochastic integral  $\int_0^t B(s) dC(s)$ , where  $B$  denotes a Brownian motion, so that  $B \in \mathcal{L}$ , and  $C$  is a compound Poisson processes with jumps  $\{Y_k\}$  that have finite mean and occur at the jump times  $\{T_k\}$  of a Poisson process  $N(t)$ , so that  $C$  is a semimartingale. Considerations as in Example 4.4 give  $\int_0^t B(s) dC(s) = \sum_{k=1}^{N(t)} B(T_k) \Delta C(T_k) = \sum_{k=1}^{N(t)} B(T_k) Y_k$  so that the jumps of  $\int_0^t B(s) dC(s)$  coincide with those of  $B(s) C(s)$ ,  $s \in (0, t]$ . Also, the Itô integral  $\int_0^t B(s) dC(s)$  coincides with the path-by-path Riemann-Stieltjes integral.  $\diamond$

The observations in Example 4.14 are consistent with results in [6] related to Itô integrals  $\int_0^t X(s) dY(s)$  with integrands  $X \in \mathcal{L}$  and semimartingale integrators  $Y$ . It is shown in [6] that the jumps of  $\int_0^t X(s) dY(s)$  are indistinguishable from the process  $X(s) \Delta Y(s)$ ,  $s \in (0, t]$  (Theorem 13, p. 53) and that  $\int_0^t X(s) dY(s)$  is indistinguishable from its definition as a path-by-path Riemann-Stieltjes integral ([6] Theorem 17, p. 54).

The following theorems show that stochastic integrals preserve an essential property of their integrators  $Y$  for integrands  $X \in \mathcal{L}$ .

**Theorem 4.9** *The process  $Z_1(t) = \int_0^t X(s) dY(s)$  is a semimartingale. Moreover, if  $G \in \mathcal{L}$ , then  $Z_2(t) = \int_0^t G(s) dZ_1(s) = \int_0^t G(s) X(s) dY(s)$  is also a semimartingale defined by a stochastic integral with integrand  $GX$  and integrator  $Y$  ([6], Theorem 19, p. 55). The latter property is referred to as associativity.*

An alternative statement of Theorem 4.9 is that the coordinates of the  $\mathbb{R}^2$ -valued process  $Z(t) = (Z_1(t), Z_2(t))$  defined by

$$dZ(t) = \begin{bmatrix} X(t) \\ G(t)X(t) \end{bmatrix} dY(t)$$

are semimartingales. Note that  $Z(t)$  can be interpreted as the state of a dynamic system subjected to semimartingale noise.

**Theorem 4.10** *If  $X \in \mathcal{L}$  and  $Y$  is a locally square integrable local martingale, then  $\int_0^t X(s) dY(s)$  is a locally square integrable local martingale ([6], Theorem 20, p. 56). If  $X \in \mathcal{L}$  and  $Y$  is a local martingale, then  $\int_0^t X(s) dY(s)$  is a local martingale ([6], Theorem 17, p. 106).*

The statement in this theorem is known as the preservation property for stochastic integrals. A property is said to hold locally for a process  $Y(t)$  if there exists an increasing sequence of stopping times  $T_1 < T_2 < \dots$  such that  $T_n \rightarrow \infty$  a.s. and  $Y(t \wedge T_n)$  has this property for each  $n \geq 1$ . If, in addition,  $Y(t \wedge T_n) \in L^2$ , then  $Y(t)$  is called locally square integrable local martingale.

*Example 4.15* The stochastic integral  $I(X)(t) = \int_0^t X(s) dB(s)$ ,  $t \geq 0$ , in Sect. 4.4.2 is a local square integrable local martingale since its integrator  $B$  is a square integrable martingale so that it is also a square integrable local martingale.  $\diamond$

## 4.7 Quadratic Variation and Covariation Processes

Let  $X$  and  $Y$  be semimartingales with the property  $X(0-) = Y(0-) = 0$  so that their jumps at  $t = 0$  are  $\Delta X(0) = X(0)$  and  $\Delta Y(0) = Y(0)$ , where  $\Delta X(t) = X(t) - X(t-) = X(t) - X_-(t)$  and  $X(t-) = X_-(t) = \lim_{s \uparrow t} X(s)$ . We define quadratic variation and covariation processes for semimartingales. These processes are essential for both developing and using Itô's formula.

**Definition 4.5** The quadratic variation process of  $X$ , denoted by  $[X, X]$  or  $[X]$ , is given by

$$[X, X](t) = X(t)^2 - 2 \int_0^t X(s-) dX(s). \quad (4.27)$$

The stochastic integral  $\int_0^t X(s-) dX(s)$  is defined since  $X_- \in \mathcal{L}$  and  $X \in \mathcal{D}$  ([6], Sects. 2.4 and 2.5). Note also that (4.27) holds at  $t = 0$  since the left and the right sides of this equation are  $[X, X](0) = (\Delta X(0))^2 = X(0)^2$  and  $X(0)^2 - 2X(0-) \Delta X(0) = X(0)^2$ .

*Example 4.16* The quadratic variation of Brownian motion is  $[B, B](t) = [B](t) = t$ . This follows from the definition of the quadratic variation process and the expression of  $\int B dB$  in Example 4.1. See also Example 3.50.  $\diamond$

*Example 4.17* The quadratic variation of a Poisson process  $N(t)$  is  $[N, N](t) = N(t)$  by (4.27) and the expression of  $\int N_- dN$  in Example 4.4. See also Example 3.52.  $\diamond$

**Theorem 4.11** The process  $[X, X]$  in (4.27) is adapted with increasing samples in  $\mathcal{D}$ , has the properties  $[X, X](0) = X(0)^2$  and  $\Delta[X, X](t) = [X, X](t) - [X, X](t-) = (\Delta X(t))^2$ , and can be obtained for each  $t > 0$  and  $s \in [0, t]$  from

$$X(0)^2 + \sum_{k=1}^n (X(t_k \wedge s) - X(t_{k-1} \wedge s))^2 \xrightarrow{\text{ucp}} [X, X](s), \quad (4.28)$$

where  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 < t_1 < \dots < t_n = t$ , is a sequence of partitions with  $\Delta(p_n) = \max_{1 \leq k \leq n} (t_k - t_{k-1}) \rightarrow 0$  as  $n \rightarrow \infty$ .

*Proof* For proof see [6] (Theorem 22, p. 59). A sequence of processes  $H_n$ ,  $n = 1, 2, \dots$ , converges to a process  $H$  uniformly on compacts in probability (ucp) if, for each  $t > 0$ ,  $\sup_{0 \leq s \leq t} |H_n(s) - H(s)| \rightarrow 0$  as  $n \rightarrow \infty$  in probability. We only note that the definitions of  $[X, X]$  and of its jumps at an arbitrary time  $t$  give  $\Delta[X, X](t) = X(t)^2 - 2 \int_0^t X(s-) dX(s) - X(t-)^2 + 2 \int_0^{t-} X(s-) dX(s) = \Delta X(t) (X(t) + X(t-)) - 2X(t-) \Delta X(t) = \Delta X(t) (X(t) - X(t-)) = (\Delta X(t))^2$ .  $\blacktriangle$

**Theorem 4.12** *If  $X$  is an  $\mathcal{F}_t$ -square integrable martingale, then  $X^2 - [X, X]$  is an  $\mathcal{F}_t$ -martingale.*

*Proof* The process has finite expectation since  $X$  is square integrable and  $[X, X] \in L^1$  ([2], Proposition 3.4). It is  $\mathcal{F}_t$ -adapted since  $X$  is a martingale and  $[X, X](t)$  depends on  $X(s)$ ,  $s \leq t$ . For  $t \geq s$ ,

$$\begin{aligned} E[X(t)^2 - X(s)^2 \mid \mathcal{F}_s] &= E[(X(t) - X(s))^2 + 2X(t)X(s) - 2X(s)^2 \mid \mathcal{F}_s] \\ &= E[(X(t) - X(s))^2 \mid \mathcal{F}_s] + 2E[X(t)X(s) \mid \mathcal{F}_s] - 2E[X(s)^2 \mid \mathcal{F}_s] \\ &= E[(X(t) - X(s))^2 \mid \mathcal{F}_s], \end{aligned}$$

which gives

$$\begin{aligned} E\left[\sum_{k=1}^n (X(t_k) - X(t_{k-1}))^2 \mid \mathcal{F}_s\right] &= \sum_{k=1}^n E\{E[(X(t_k) - X(t_{k-1}))^2 \mid \mathcal{F}_{t_{k-1}}] \mid \mathcal{F}_s\} \\ &= \sum_{k=1}^n E\{E[X(t_k)^2 - X(t_{k-1})^2 \mid \mathcal{F}_{t_{k-1}}] \mid \mathcal{F}_s\} = \sum_{k=1}^n E[(X(t_k)^2 - X(t_{k-1})^2) \mid \mathcal{F}_s] \\ &= E\left[\sum_{k=1}^n (X(t_k)^2 - X(t_{k-1})^2) \mid \mathcal{F}_s\right] = E[X(t)^2 - X(s)^2 \mid \mathcal{F}_s] \end{aligned}$$

for any partition  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $s = t_0 \leq t_1 \leq \dots \leq t_n = t$ , of the interval  $[s, t]$ . The left side of the above equalities converges in  $L^1$  to  $[X, X](t) - [X, X](s)$  as  $\Delta(p_n) \rightarrow 0$  ([2], Proposition 3.4) so that  $E[X(t)^2 - [X, X](t) \mid \mathcal{F}_s] = E[X(s)^2 - [X, X](s) \mid \mathcal{F}_s] = X(s)^2 - [X, X](s)$  a.s. showing that  $X^2 - [X, X]$  is a martingale. It can also be shown that, if  $X$  is a local square integrable martingale,  $X^2 - [X, X]$  is a local martingale ([2], Proposition 6.1).  $\blacktriangle$

Let  $X(t)$  be a square integrable martingale. Then  $X(t)^2 - \langle X \rangle(t)$  and  $X(t)^2 - [X, X](t)$  are martingales, where  $\langle X \rangle$  denotes the compensator in the Doob-Meyer decomposition. Since the collection of martingale defines a linear space, then  $\langle X \rangle - [X, X]$  is also a martingale. We have seen that  $\langle X \rangle$  and  $[X, X]$  are indistinguishable for continuous martingale. Generally,  $\langle X \rangle$  and  $[X, X]$  differ, as illustrated by the following example.

**Example 4.18** Let  $X(t) = N(t) - t$ ,  $t \geq 0$ , where  $N$  be a Poisson process with unit intensity. The compensator of  $X(t)^2$  is  $\langle X \rangle(t) = t$  (Example 4.9), while the quadratic variation process of  $X(t)$  is  $[X, X](t) = N(t)$  (Example 3.52).  $\diamond$

**Example 4.19** Let  $C(t) = \sum_{k=1}^{N(t)} Y_k$  be a compound Poisson process, where  $N$  is a Poisson process with intensity  $\lambda > 0$  and  $\{Y_k\}$  are iid random variables. The stochastic integral  $\int C_- dC$  is  $\int_0^t C(s-) dC(s) = (C(t)^2 - \sum_{k=1}^{N(t)} Y_k^2)/2$ . If  $Y_k = 1$  a.s., then  $\int_0^t C(s-) dC(s) = \int_0^t N(s-) dN(s) = (N(t)^2 - N(t))/2$ , in agreement with Example 4.4.  $\diamond$



*Proof* Theorem 4.11 implies  $[C, C](t) = \sum_{k=1}^{N(t)} Y_k^2$  since the increments  $\Delta C(s) = C(s) - C(s-)$  are non-zero and equal to  $Y_k$  at the jump times  $T_k$  of  $N(t)$ . The definition of quadratic variation in (4.27) gives  $\sum_{k=1}^{N(t)} Y_k^2 = C(t)^2 - 2 \int_0^t C(s-) dC(s)$ , which yields the stated equality.  $\blacktriangle$

*Example 4.20* Let  $C(t) = \sum_{k=1}^{N(t)} Y_k$  be a compound Poisson process as in the previous example. Then  $\Delta[C, C](t) = Y_{N(t)}^2 1(\Delta N(t) = 1)$ , where  $\Delta N(t) = N(t) - N(t-)$ .  $\diamond$

*Proof* If  $C$  has a jump at a time  $t \geq 0$ , then  $\Delta C(t) = Y_{N(t)} 1(\Delta N(t) = 1)$ . The expression of  $\Delta[C, C](t)$  results from Theorem 4.11.  $\blacktriangle$

**Definition 4.6** The quadratic covariation of  $X$  and  $Y$  is

$$[X, Y](t) = X(t) Y(t) - \int_0^t X(s-) dY(s) - \int_0^t Y(s-) dX(s). \quad (4.29)$$

The definition coincides with (4.27) for  $X = Y$ . It is meaningful since  $X$  and  $Y$  are semimartingales so that the stochastic integrals  $\int X_- dY$  and  $\int Y_- dX$  are defined ([6], Sects. 2.4 and 2.5). That (4.29) holds at  $t = 0$  can be shown by using arguments as for (4.27).

**Theorem 4.13** For each  $t > 0$  and partitions  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 < t_1 < \dots < t_n = t$ , such that  $\Delta(p_n) \rightarrow 0$  as  $n \rightarrow \infty$ , we have

$$\begin{aligned} X(0) Y(0) + \sum_{k=1}^n (X(t_k \wedge s) - X(t_{k-1} \wedge s)) (Y(t_k \wedge s) - Y(t_{k-1} \wedge s)) \\ \xrightarrow{\text{ucp}} [X, Y](s), \quad s \in [0, t], \end{aligned} \quad (4.30)$$

$[X, Y](0) = X(0) Y(0)$ , and  $\Delta[X, Y](t) = \Delta X(t) \Delta Y(t)$  ([6], Theorem 23, p. 61).

**Theorem 4.14** The equalities

$$\begin{aligned} [X, Y] &= \frac{1}{2} ([X + Y, X + Y] - [X, X] - [Y, Y]) \quad \text{and} \\ \int_{0+}^t X(s-) dY(s) &= X(t) Y(t) - \int_{0+}^t Y(s-) dX(s) - [X, Y](t) \end{aligned} \quad (4.31)$$

hold, and are referred to polarization identity and integration by parts, respectively.

*Proof* Since the collection of semimartingales defines a linear space, the quadratic variation process  $[X + Y, X + Y]$  is defined. By properties of the stochastic integral and the definitions of the quadratic variation and covariation processes, we have  $[X + Y, X + Y] = [X, X] + [Y, Y] + 2[X, Y]$ , which yields the polarization identity.

The integration by parts formula is a direct consequence of (4.29). If  $[X, Y] = 0$ , (4.31) coincides with the integration by parts formula of the classical calculus. The notation  $\int_{0+}^t$  means that the integrals are performed in  $(0, t]$ . The integration by

parts formula can also be used in  $[0, t]$  in which case is written as  $\int_0^t X(s-) dY(s) = X(t)Y(t) - \int_0^t Y(s-) dX(s) - [X, Y](t)$ . That the formula holds at  $t = 0$  follows by direct calculations.  $\blacktriangle$

*Example 4.21* The quadratic covariation  $[B_1, B_2]$  of two independent Brownian motions  $B_1$  and  $B_2$  is zero.  $\diamond$

*Proof* The polarization identity gives  $[B_1, B_2](t) = ([\sqrt{2}B, \sqrt{2}B](t) - [B_1, B_1](t) - [B_2, B_2](t))/2 = 0$ , where  $\sqrt{2}B(t)$  and  $B_1(t) + B_2(t)$  are versions and  $B$  denotes a Brownian motion.  $\blacktriangle$

**Definition 4.7** The path-by-path continuous part  $[X, X]^c$  of  $[X, X]$  is defined by

$$\begin{aligned} [X, X](t) &= [X, X]^c(t) + X(0)^2 + \sum_{0 < s \leq t} (\Delta X(s))^2 \\ &= [X, X]^c(t) + \sum_{0 \leq s \leq t} (\Delta X(s))^2. \end{aligned} \quad (4.32)$$

If  $[X, X]^c(t) = 0$ , then  $X$  is said to be a quadratic pure jump semimartingale. The quadratic variation of a quadratic pure jump semimartingale  $X$  is  $[X, X](t) = \sum_{0 \leq s \leq t} (\Delta X(s))^2$ . It can be shown that any semimartingale  $X$  has a unique continuous local martingale part  $X^c$  and that  $[X^c, X^c] = [X, X]^c$  ([6], p. 63).

**Theorem 4.15** If  $X$  is a quadratic pure jump semimartingale and  $Y$  is an arbitrary semimartingale, then ([6], Theorem 28, p. 68)

$$[X, Y](t) = X(0)Y(0) + \sum_{0 < s \leq t} \Delta X(s) \Delta Y(s). \quad (4.33)$$

*Example 4.22* The quadratic covariation  $[B, C]$  of a Brownian motion  $B$  and the compound Poisson process  $C$  is zero, so that

$$[B + C, B + C](t) = [B, B](t) + [C, C](t) = t + \sum_{k=1}^{N(t)} Y_k^2.$$

The quadratic variation of  $B + C$  varies linearly in time between the jumps of the pure jump semimartingale  $C(t)$ .  $\diamond$

*Proof* The polarization identity (4.31) with  $(B, C)$  in place of  $(X, Y)$  gives  $[B, C] = ([B + C, B + C] - [B, B] - [C, C])/2$ . We have already found the expressions for the quadratic variation processes  $[B, B]$  and  $[C, C]$ . The quadratic covariation process  $[B, C]$  is zero by (4.33) and properties of  $(B, C)$ .  $\blacktriangle$

Note that  $X^c = B$  for the semimartingale  $X = B + C$  so that  $[X, X]^c(t) = t$ ,  $[X, X](t) = t + \sum_{k=1}^{N(t)} Y_k^2$ , and  $[X, X](t) - [X, X]^c(t) = \sum_{k=1}^{N(t)} Y_k^2 = \sum_{0 \leq s \leq t} (\Delta C(s))^2$ , where  $B$  and  $C$  denote a Brownian motion and a compound Poisson process, respectively. The process  $[X, X](t) - [X, X]^c(t)$  is given by the sum of the

squares of all jumps of  $C$  in  $[0, t]$ . This result holds for any semimartingale  $X$  since semimartingales can have only jump discontinuities and the number of jumps is at most countable. We also note that  $[X, X](0) = X(0)^2$  and  $[X, X]^c(0) = 0$  since the increment  $\Delta X(0) = X(0) - X(0-)$  of  $X$  at  $t = 0$  is  $X(0)$ .

**Theorem 4.16** *The quadratic variation and covariation processes  $[X, X]$  in (4.27) and  $[X, Y]$  in (4.29) are semimartingales. Also, if  $X$  has a.s. continuous sample paths of finite variation on compacts, then  $[X, X](t) = X(0)^2$ .*

*Proof* Since  $[X, X]$  is an adapted process with increasing samples in  $\mathcal{D}$ , the process  $[X, X]$  is a semimartingale. The polarization identity and the fact that the collection of martingales is a linear space show that the quadratic covariation process  $[X, Y]$  is also a semimartingale.

We also note that the product  $X(t)Y(t)$  of two semimartingales is a semimartingale since stochastic integrals and quadratic covariation processes are semimartingales and  $X(t)Y(t)$  is a linear function of these processes by (4.29).

For the sequence of partition  $p_n$  in (4.30)

$$\sum_{k=1}^n (X(t_k) - X(t_{k-1}))^2 \leq \sup_k |X(t_k) - X(t_{k-1})| \sum_{k=1}^n |X(t_k) - X(t_{k-1})| \rightarrow 0$$

as  $n \rightarrow \infty$  since  $\sup_k |X(t_k) - X(t_{k-1})| \rightarrow 0$  and  $\sum_{k=1}^n |X(t_k) - X(t_{k-1})| < \infty$  by the continuity and the finite variation of the samples of  $X$ , respectively, so that  $[X, X](t) = X(0)^2$ .  $\blacktriangle$

**Example 4.23** If  $X$  is a process with continuous samples of finite variation on compacts, then  $\int_0^t X(s-) dX(s) = \int_0^t X(s) dX(s) = (X(t)^2 - X(0)^2)/2$  and  $[X, X](t) = X(0)^2$ .  $\diamond$

*Proof* We have  $[X, X](t) = X(0)^2$  by Theorem 4.16 so that  $\int_0^t X_- dX$  results from (4.27). For example, take  $X(t) = A \cos(\nu t)$ , where  $A$  is a real-valued random variable and  $\nu > 0$  is a real number. Then

$$\begin{aligned} \int_0^t X(s-) dX(s) &= \int_0^t X(s) dX(s) = \int_0^t (A \cos(\nu s)) d(A \cos(\nu s)) \\ &= (A^2/2)[(\cos(\nu t))^2 - 1] = X(t)^2/2 - A^2/2 \end{aligned}$$

so that  $[X, X](t) = X(t)^2 - 2(X(t)^2/2 - A^2/2) = A^2 = X(0)^2$ .  $\blacktriangle$

## 4.8 Exercises

**Exercise 4.1** Complete the calculations in Example 4.1 showing that  $J_{B,n}(B)$  converges in m.s. to  $(B(t)^2 - t)/2$ .

**Exercise 4.2** Calculate the distribution of the random variable  $\int_0^1 B(t) dB(t)$ , where  $B$  is a Brownian motion and  $\int B dB$  denotes a stochastic integral.

**Exercise 4.3** Find the Itô integral  $\int_0^t C(s-) dC(s)$  and the Stratonovich integral  $\int_0^t C(s-) \circ dC(s)$  by using arguments as in Example 4.4, where  $C(t)$  is a compound Poisson process.

**Exercise 4.4** Show that  $\mathcal{H}_0^2[0, \tau]$  in (4.10) is a linear space and that the stochastic integral  $I(X)$  in (4.12) is linear in  $X$ .

**Exercise 4.5** Calculate the mean and covariance functions of the process  $X(t) = \int_0^t g(s) dB(s)$  defined as a path by path Riemann-Stieltjes integral, where  $g$  is a real-valued function of bounded variation and  $B$  denotes a Brownian motion.

**Exercise 4.6** Find the Itô integrals  $\int_0^t C(s-) dB(s)$  and  $\int_0^t B(s) dC(s)$  by following the approach in Examples 4.1 and 4.4.

**Exercise 4.7** Calculate the mean and variance of the random variables defined by the Itô integrals  $\int_0^1 |B(t)| dB(t)$ ,  $\int_0^1 \sqrt{t} \exp(B(t)) dB(t)$ , and  $\int_0^1 \sqrt{t} \sin(B(t)) dB(t)$ , where  $B$  is a Brownian motion and the integrals are stochastic integrals.

**Exercise 4.8** Show that  $X(t) = \int_0^t g(s) dB(s)$ ,  $0 \leq t \leq \tau$ , is a Gaussian process with mean 0 and covariance function  $E[X(s)X(t)] = \int_0^{s \wedge t} g(u)^2 du$ , where  $g: \mathbb{R} \rightarrow \mathbb{R}$  is of bounded variation.

**Exercise 4.9** Use Itô's isometry to calculate the variance of  $\int_0^t |B(s)|^{1/2} dB(s)$  and  $\int_0^t (B(s) + s)^2 dB(s)$ .

*Hint* Itô's isometry gives  $E[(\int_0^t |B(s)|^{1/2} dB(s))^2] = E[\int_0^t |B(s)| ds]$  so that the variance of the first integral can be calculated from  $\int_0^t E[|B(s)|] ds$  by Fubini's theorem.

**Exercise 4.10** Complete the calculations in Example 4.9 showing that the compensator of  $N(t) - \lambda t$  is  $\lambda t$ .

**Exercise 4.11** The process  $M(t) = B(t)^2 - t$  is a martingale with respect to the filtration generated by the Brownian motion  $B$ . Show that the compensator of  $M(t)$  in Doob-Mayer's decomposition is  $\langle M \rangle(t) = 4 \int_0^t B(s)^2 ds$ .

**Exercise 4.12** Show that  $M(t) = \exp(\lambda B(t) - \lambda^2 t/2)$ ,  $\lambda > 0$ , is a martingale and that the compensator of  $M(t)^2$  is  $\langle M \rangle(t) = \lambda^2 \int_0^t \exp(2\lambda B(s) - \lambda^2 s) ds$ .

**Exercise 4.13** Show that  $C(t)^2$  is a semimartingale, where  $C(t)$  denotes a compound Poisson process as in Example 4.12.

**Exercise 4.14** Find the quadratic variation of the semimartingale  $Y$  in (4.26).

**Exercise 4.15** Let  $\tilde{N}(t)$  be a compensated Poisson process that is independent of a Brownian motion  $B$ . Find the variance of  $\int_a^b B(s) d\tilde{N}(s)$  and  $\int_a^b \exp(B(s)) d\tilde{N}(s)$ .

**Exercise 4.16** Find and plot the quadratic variation for the approximate representation  $\tilde{L}_{\alpha,a}(t)$  of an  $\alpha$ -stable process  $L_\alpha(t)$  defined by (3.54) for several values of  $\alpha \in (1, 2]$ .

## References

1. Apostol TM (1974) Mathematical analysis. Addison-Wesley Publishing Company, Reading
2. Ethier SN, Kurtz TG (1986) Markov processes: Characterization and convergence. Wiley, New York
3. Grigoriu M (1995) Applied non-Gaussian processes: Examples, theory, simulation, linear random vibration, and MATLAB solutions. Prentice Hall, Englewoods Cliffs
4. Kuo H-H Introduction to stochastic integration. Springer, New York.
5. Mikosch T (1998) Elementary stochastic calculus. World Scientific, New Jersey
6. Protter P (1990) Stochastic integration and differential equations. Springer, New York
7. Steele JM (2001) Stochastic calculus and financial applications. Springer, New York
8. Taylor SJ (1972) Exact asymptotic estimates of Brownian path variation. Duke Math J 39: 219–241

# Chapter 5

## Itô's Formula and Applications

### 5.1 Introduction

The Itô formula extends the change of variable formula of the classical calculus to stochastic integrals of the type examined in the previous chapter, and constitutes an essential tool for solving stochastic problems encountered in physics and engineering. The classical change of variables formula,

$$g(h(t)) - g(h(0)) = \int_{h(0)}^{h(t)} g'(u) du = \int_0^t g'(h(s)) dh(s), \quad (5.1)$$

gives the increment of a deterministic real-valued function  $t \mapsto g(h(t))$  in an interval  $[0, t]$ . The differential form of this formula is

$$\frac{d}{dt} [g(h(t))] = g'(h(t)) h'(t) \quad \text{or} \quad d[g(h(t))] = g'(h(t)) dh(t), \quad (5.2)$$

where  $g'$  and  $h'$  denote the first derivatives of functions  $g$  and  $h$ . The Itô formula extends the rules of classical calculus to the case in which the deterministic function  $h$  is replaced with a semimartingale  $X$ .

The change of variables formula in (5.1) gives  $B(t)^2/2 = \int_0^t B(s) dB(s)$  for  $g(y) = y^2/2$  and  $h$  replaced by a Brownian motion process  $B$ . The result is in disagreement with the Itô integral  $\int_0^t B(s) dB(s) = B(t)^2/2 - t/2$  in (4.6) but is consistent with the Stratonovich integral  $\int_0^t B(s) \circ dB(s) = B(t)^2/2$  given by (4.8).

### 5.2 Itô's Formula for $\mathbb{R}$ -Valued Semimartingales

We establish the change of variable formula for real-valued functions  $t \mapsto g(X(t))$ , where  $g \in C^2(\mathbb{R})$  and  $X$  denotes a real-valued semimartingale. The formula, referred to as Itô's formula, is established for continuous semimartingales and is extended

subsequently to arbitrary semimartingales. Examples are used to illustrate the application and usefulness of Itô's formula.

### 5.2.1 Continuous Semimartingales

This section outlines the essential steps of the proof of Itô's formula. Additional technical details can be found in [15] (Theorem 32, p. 71). A heuristic derivation of the Itô formula is in [11] (Theorem 6.7.1).

**Theorem 5.1** *If  $X$  is a continuous semimartingale and  $g \in C^2(\mathbb{R})$ , then  $g(X)$  is a continuous semimartingale and, for all  $t \geq 0$ , the integral and differential forms,*

$$\begin{aligned} g(X(t)) - g(X(0)) &= \int_0^t g'(X(s)) dX(s) + \frac{1}{2} \int_0^t g''(X(s)) d[X, X](s) \quad \text{and} \\ dg(X(t)) &= g'(X(t)) dX(t) + \frac{1}{2} g''(X(t)) d[X, X](t), \end{aligned} \quad (5.3)$$

of Itô's formula hold with probability 1.

*Proof* If (5.3) holds,  $g(X)$  is a semimartingale since (1) the processes  $g'(X)$  and  $g''(X)$  are adapted as memoryless transformations of the adapted process  $X$ , (2)  $g'(X)$  and  $g''(X)$  have continuous samples since  $g \in C^2(\mathbb{R})$  and  $X$  is continuous by assumption, (3) the integrals  $\int_0^t g'(X(s)) dX(s)$  and  $\int_0^t g''(X(s)) d[X, X](s)$  are semimartingales by a preservation property of the stochastic integral ([15], Theorem 20, p. 56 and Theorem 17, p. 106, and Theorem 4.10 in this book), and (4) sums of semimartingales are semimartingales.

The Taylor formula cannot be applied directly to  $g(X)$  since  $X$  may not take values in a bounded interval, but it can be used for  $X$  stopped at time  $T_a = \inf\{t \geq 0 : |X(t)| \geq a\}$ ,  $0 < a < \infty$ . Since  $a$  is arbitrary, results established for  $X$  stopped at  $T_a$  hold for  $a \rightarrow \infty$ , that is, for  $X$ .

The Taylor formula shows that the increment  $g(x+h) - g(x)$  of function  $g$  in  $[x, x+h]$ ,  $h > 0$ , has the form

$$g(x+h) - g(x) = hg'(x) + \frac{h^2}{2} g''(x) + r(x, h), \quad h \in \mathbb{R}, \quad (5.4)$$

where  $|r(x, h)| \leq h^2 \alpha(|h|)$ ,  $\alpha : [0, \infty) \rightarrow [0, \infty)$  is increasing, and  $\lim_{u \downarrow 0} \alpha(u) = 0$ . Fix  $t > 0$  and consider a sequence of partitions  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 \leq t_1 \leq \dots \leq t_n = t$ , of  $[0, t]$  such that  $\Delta(p_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Then

$$g(X(t)) - g(X(0)) = \sum_{k=1}^n (g(X(t_k)) - g(X(t_{k-1}))) = S_1 + S_2 + S_3, \quad (5.5)$$

where

$$\begin{aligned} S_1 &= \sum_{k=1}^n g'(X(t_{k-1})) (X(t_k) - X(t_{k-1})) \\ S_2 &= \frac{1}{2} \sum_{k=1}^n g''(X(t_{k-1})) (X(t_k) - X(t_{k-1}))^2 \\ S_3 &= \sum_{k=1}^n r(X(t_{k-1}), X(t_k) - X(t_{k-1})) \end{aligned}$$

correspond to the three terms of the Taylor formula in (5.4) applied to the increments  $g(X(t_k)) - g(X(t_{k-1}))$ . The sums  $S_1$  and  $S_2$  converge in probability to the Itô integrals  $\int_0^t g'(X(s)) dX(s)$  and  $(1/2) \int_0^t g''(X(s)) d[X, X](s)$  as  $n \rightarrow \infty$ , respectively ([15], Theorem 21, p. 57, and Theorem 30, p. 69, and Sects. 4.4, 4.5, and 4.6 in this book). The absolute value of  $S_3$  can be bounded by

$$\begin{aligned} |S_3| &= \left| \sum_{k=1}^n r(X(t_{k-1}), X(t_k) - X(t_{k-1})) \right| \\ &\leq \max_{1 \leq k \leq n} \alpha(|X(t_k) - X(t_{k-1})|) \sum_{k=1}^n (X(t_k) - X(t_{k-1}))^2 \end{aligned}$$

for each  $n$ . Since  $X$  has continuous samples, the function  $s \mapsto X(s, \omega)$  is uniformly continuous in  $[0, t]$  for almost all  $\omega$ , so that  $\max_k |X(t_k) - X(t_{k-1})| \rightarrow 0$  a.s. as  $n \rightarrow \infty$  implying  $\max_k \alpha(|X(t_k) - X(t_{k-1})|) \rightarrow 0$  a.s. We conclude that  $S_3 \rightarrow 0$  in probability as  $n \rightarrow \infty$  since  $\max_{1 \leq k \leq m_n} \alpha(|X(t_k) - X(t_{k-1})|) \rightarrow 0$  a.s. and  $\sum_k (X(t_k) - X(t_{k-1}))^2 \xrightarrow{\text{ucp}} [X, X](t)$  (Theorem 4.11).

In summary, we have shown that for each  $t \geq 0$  the sequences  $S_1$ ,  $S_2$ , and  $S_3$  converge in probability to  $\int_0^t g'(X(s)) dX(s)$ ,  $(1/2) \int_0^t g''(X(s)) d[X, X](s)$ , and zero, respectively, as  $n \rightarrow \infty$ . Hence, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left( \left| S_1 + S_2 + S_3 - \int_0^t g'(X(s)) dX(s) \right. \right. \\ \left. \left. - \frac{1}{2} \int_0^t g''(X(s)) d[X, X](s) \right| > \varepsilon \right) = 0 \end{aligned}$$

for any  $\varepsilon > 0$ , so that the first equality in (5.3) holds with probability 1.  $\blacktriangle$

*Example 5.1* The Itô formula in (5.3) applied to the mapping  $B \mapsto g(B) = B^n$  gives

$$B(t)^n = n \int_0^t B(s)^{n-1} dB(s) + \frac{n(n-1)}{2} \int_0^t B(s)^{n-2} ds, \quad (5.6)$$



where  $n \geq 1$  is an integer. For the special case  $n = 2$  we have  $B(t)^2 = 2 \int_0^t B(s) dB(s) + t$  in agreement with our calculations in Example 4.1. Note that  $B(t)^n$  is a semimartingale and that (5.6) provides a recurrence formula for calculating the Itô integrals  $\int_0^t B^n dB$ .  $\diamond$

*Proof* The Itô formula in (5.3) can be used since the mapping  $B \mapsto B^n$  is infinitely differentiable and  $B$  is a continuous square integrable martingale, so that it is a continuous semimartingale. That  $g(B) = B^2$  is a semimartingale also follows from the representation  $B(t)^2 = A(t) + M(t)$ , where  $A(t) = t$  is an adapted continuous process with  $A(0) = 0$  and paths of finite variation on compacts and  $M(t) = B(t)^2 - t$  is a square integrable martingale starting at zero.  $\blacktriangle$

### 5.2.2 Arbitrary Semimartingales

The extension of the Itô formula to arbitrary semimartingales is based on the fact that semimartingales have at most countable numbers of jumps in bounded time intervals ([15], Sect. 1.1) and continuous samples between consecutive jumps.

**Theorem 5.2** *If  $X$  is a semimartingale and  $g \in C^2(\mathbb{R})$ , then  $g(X)$  is a semimartingale and, for all  $t \geq 0$ , the integral form,*

$$\begin{aligned} g(X(t)) - g(X(0)) &= \int_{0+}^t g'(X(s-)) dX(s) + \frac{1}{2} \int_{0+}^t g''(X(s-)) d[X, X]^c(s) \\ &\quad + \sum_{0 < s \leq t} [g(X(s)) - g(X(s-)) - g'(X(s-)) \Delta X(s)], \end{aligned} \quad (5.7)$$

*of Itô's formula holds with probability 1.*

The Itô formula can also be given in the form

$$\begin{aligned} g(X(t)) - g(X(0)) &= \int_{0+}^t g'(X(s-)) dX(s) + \frac{1}{2} \int_{0+}^t g''(X(s-)) d[X, X](s) \\ &\quad + \sum_{0 < s \leq t} \left[ g(X(s)) - g(X(s-)) - g'(X(s-)) \Delta X(s) - \frac{1}{2} g''(X(s-)) (\Delta X(s))^2 \right], \end{aligned} \quad (5.8)$$

by using the relationship between  $[X, X]$  and  $[X, X]^c$  (Definition 4.7). The Itô formulas in (5.7) and (5.8) do not include the jumps of  $X$  at 0. To include these jumps, it is sufficient to write  $\int_0^t$  and  $\sum_{0 \leq s \leq t}$  in place of  $\int_{0+}^t$  and  $\sum_{0 < s \leq t}$ .

*Proof* If (5.7) holds,  $g(X)$  is a semimartingale since (1) the processes  $g'(X)$  and  $g''(X)$  are adapted as memoryless transformations of the adapted process  $X$ , (2)  $g'(X)$  and  $g''(X)$  are adapted processes that have right continuous samples with left limits, since, for example,  $\lim_{s \downarrow t} g'(X(s)) = g'(\lim_{s \downarrow t} X(s)) = g'(X(t))$  and

$\lim_{s \uparrow t} g'(X(s)) = g'(\lim_{s \uparrow t} X(s)) = g'(X(t-))$ , (3) the integrals  $\int_0^t g'(X(s)) dX(s)$  and  $\int_0^t g''(X(s)) d[X, X]^c(s)$  are semimartingales by a preservation property of the stochastic integral ([15], Theorem 20, p. 56, and Theorem 17, p. 106), (4) the summation in (5.7) is a semimartingale (Exercise 5.2), and (5) sums of semimartingales are semimartingales.

Let  $[t_{i-1}, t_i]$  denote a time interval between two consecutive jumps of  $X$  in a time interval  $(0, t]$ . We have

$$g(X(t_i)) - g(X(t_{i-1})) = [g(X(t_i)) - g(X(t_i-))] + [g(X(t_i-) - g(X(t_{i-1}))],$$

where the first and the second terms on the right side of this equation correspond to the jump of  $X$  at  $t_i$  and the change of  $X$  in the continuity interval  $[t_{i-1}, t_i]$ . The contribution of terms associated with the jumps and the continuity intervals of  $X$  in  $(0, t]$  are  $\sum_{0 < s \leq t} [g(X(s)) - g(X(s-))]$  and (Theorem 5.1)

$$g(X(t_i-) - g(X(t_{i-1})) \simeq g'(X(t_{i-1})) (X(t_i-) - X(t_{i-1})) + \frac{1}{2} g''(X(t_{i-1})) (X(t_i-) - X(t_{i-1}))^2. \quad (5.9)$$

The first term on the right side of (5.9) has the representation  $g'(X(t_{i-1}))(X(t_i-) - X(t_{i-1})) = g'(X(t_{i-1})) [X(t_i) - X(t_{i-1}))] - g'(X(t_{i-1})) [X(t_i) - X(t_i-)]$ , so that its contribution in  $(0, t]$  becomes

$$\begin{aligned} \sum_i g'(X(t_{i-1})) [X(t_i) - X(t_{i-1}))] &\rightarrow \int_{0+}^t g'(X(s-)) dX(s) \text{ in probability} \\ \sum_i g'(X(t_{i-1})) [X(t_i) - X(t_i-)] &= \sum_{0 < s \leq t} g'(X(s-)) [X(s) - X(s-)], \end{aligned}$$

as the mesh  $\Delta(p_n) = \max_{1 \leq k \leq n} (t_k - t_{k-1})$  of partition  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 < t_1 < \dots < t_n = t$ , of  $(0, t]$  approaches 0 (Sect. 4.5). The contribution of the second term on the right side of (5.9) in  $(0, t]$  converges to

$$\frac{1}{2} \sum_i g''(X(t_{i-1})) (X(t_i-) - X(t_{i-1}))^2 \rightarrow \frac{1}{2} \int_{0+}^t g''(s-) d[X, X]^c(s) \text{ in probability,}$$

as  $\Delta(p_n) \rightarrow 0$  by Theorem 4.11 giving the convergence  $X(0)^2 + \sum_{k=1}^n (X(t_k) - X(t_{k-1}))^2 \rightarrow [X, X](t)$  in probability. These observations yield the Itô formula in (5.7). The version of this formula in (5.8) follows from the relationship  $[X, X](t) = [X, X]^c(t) + X(0)^2 + \sum_{0 < s \leq t} (\Delta X(s))^2$  between the processes  $[X, X]$  and  $[X, X]^c$  (Definition 4.7).  $\blacktriangle$

*Example 5.2* The process  $N^2$  is a semimartingale and

$$N(t)^2 = 2 \int_{0+}^t N(s-) dN(s) + N(t), \quad (5.10)$$

where  $N$  is a Poisson process with intensity  $\lambda > 0$ .  $\diamond$

*Proof* Theorem 5.2 shows that  $N^2$  is a semimartingale. Itô's formula applied to  $g(N) = N^2$  gives

$$\begin{aligned} N(t)^2 - N(0)^2 &= \int_{0+}^t 2N(s-) dN(s) + \frac{1}{2} \int_{0+}^t 2 d[N, N]^c(s) \\ &\quad + \sum_{0 < s \leq t} [N(s)^2 - N(s-)^2 - 2N(s-) \Delta N(s)]. \end{aligned}$$

Since  $N$  is a quadratic pure jump semimartingale,  $[N, N]^c(t) = 0$  for all  $t \geq 0$  so that  $\int_{0+}^t 2 d[N, N]^c(s) = 0$ . The above summation has the alternative form

$$\sum_{i=1}^{N(t)} [N(T_i)^2 - N(T_{i-1})^2 - 2N(T_{i-1})] = \sum_{i=1}^{N(t)} [i^2 - (i-1)^2 - 2(i-1)] = N(t),$$

where  $\{T_i\}$  denote the jump times of  $N$ , which gives (5.10) since  $N(0) = 0$ .  $\blacktriangle$

*Example 5.3* Let  $N(t)$  be a Poisson process as in the previous example. The recurrence formula

$$\begin{aligned} \int_{0+}^t N(s-)^n dN(s) &= \int_{0+}^t N(s-)^{n-1} dN(s) + \frac{N(t)^{n+1}}{n+1} - \frac{N(t)^n}{n} \\ &\quad - \sum_{i=1}^{N(t)} \left[ \frac{i^{n+1} - (i-1)^{n+1} - (n+1)(i-1)^n}{n+1} - \frac{i^n - (i-1)^n - n(i-1)^{n-1}}{n} \right] \end{aligned} \quad (5.11)$$

holds, where  $n \geq 1$  is an integer. This formula gives  $\int_{0+}^t N(s-) dN(s) = (N(t)^2 - N(t))/2$  for  $n = 1$  in agreement with (5.10).  $\diamond$

*Proof* Itô's formula applied to the mapping  $N(t) \mapsto N(t)^n$  gives

$$\begin{aligned} N(t)^n &= n \int_{0+}^t N(s-)^{n-1} dN(s) + \sum_{0 < s \leq t} [N(s)^n - N(s-)^n - nN(s-)^{n-1} \Delta N(s)] \\ &= \int_{0+}^t N(s-)^{n-1} dN(s) + \sum_{i=1}^{N(t)} [N(T_i)^n - N(T_{i-1})^n - nN(T_{i-1})^{n-1}], \end{aligned}$$

where  $\{T_i\}$  denote the jump times of  $N$  and the latter summation can be written as  $\sum_{i=1}^{N(t)} [i^n - (i-1)^n - n(i-1)^{n-1}]$ . The recurrence formula in (5.11) results by subtracting the above equation from that for  $n+1$ .  $\blacktriangle$

*Example 5.4* Let  $C(t) = \sum_{i=1}^{N(t)} Y_i$  be a compound Poisson process, where  $N(t)$  is a Poisson process with intensity  $\lambda > 0$  and  $\{Y_i\}$  are iid real-valued random variables.

The process

$$C(t)^2 = 2 \int_{0+}^t C(s-) dC(s) + \sum_{i=1}^{N(t)} Y_i^2 \quad (5.12)$$

is a semimartingale.  $\diamond$

*Proof* Itô's formula in (5.7) applied to  $g(C(t)) = C(t)^2$  gives

$$\begin{aligned} C(t)^2 - C(0)^2 &= \int_{0+}^t 2 C(s-) dC(s) + \frac{1}{2} \int_{0+}^t 2 d[C, C]^c(s) \\ &\quad + \sum_{0 < s \leq t} [C(s)^2 - C(s-)^2 - 2 C(s-) \Delta C(s)]. \end{aligned} \quad (5.13)$$

Since  $C$  is a quadratic pure jump semimartingale,  $\int_{0+}^t d[C, C]^c(s) = 0$  so that

$$\begin{aligned} \sum_{0 < s \leq t} [C(s)^2 - C(s-)^2 - 2 C(s-) \Delta C(s)] &= \sum_{i=1}^{N(t)} [C(T_i)^2 - C(T_{i-1})^2 - 2 C(T_{i-1}) Y_i] \\ &= \sum_{i=1}^{N(t)} \left[ \left( \sum_{k=1}^i Y_k \right)^2 - \left( \sum_{k=1}^{i-1} Y_k \right)^2 - 2 \left( \sum_{k=1}^{i-1} Y_k \right) Y_i \right] = \sum_{i=1}^{N(t)} Y_i^2, \end{aligned}$$

where the latter equality holds since the square bracket has the form  $(Z + Y_i)^2 - Z^2 - 2ZY_i$  with  $Z = \sum_{k=1}^{i-1} Y_k$  so that it is  $Y_i^2$ . Hence, the summation in (5.13) is a compound Poisson process with jumps  $\{Y_i^2\}$  occurring at the jump times  $T_i$  of  $C$ . These considerations and  $C(0) = 0$  give (5.12). Note that (5.12) provides an alternative definition for the stochastic integral  $\int C_- dC$ .  $\blacktriangle$

*Example 5.5* Let  $C$  be the compound Poisson process in the previous example and  $g$  a real-valued function with continuous second order derivative. An alternative form of the Itô formula in (5.7) is

$$g(C(t)) - g(C(0)) = \int_{0+}^t \int_{\mathbb{R}} [g(C(s-) + y) - g(C(s-))] \mathcal{M}(ds, dy), \quad (5.14)$$

where  $\mathcal{M}(ds, dy)$  denotes a random measure giving the number of jumps of  $C$  in the rectangle  $(s, s + ds] \times (y, y + dy]$  (Definition 3.38).  $\diamond$

*Proof* Itô's formula in (5.7) applied to  $g(C(t))$  gives

$$\begin{aligned} g(C(t)) - g(C(0)) &= \int_{0+}^t g'(C(s-)) dC(s) + \frac{1}{2} \int_{0+}^t g''(C(s-)) d[C, C]^c(s) \\ &\quad + \sum_{0 < s \leq t} [g(C(s)) - g(C(s-)) - g'(C(s-)) \Delta C(s)] = \sum_{0 < s \leq t} [g(C(s)) - g(C(s-))] \end{aligned}$$

since  $\int_{0+}^t g'(C(s-)) dC(s) = \sum_{0 < s \leq t} g'(C(s-)) \Delta C(s)$  and  $[C, C]^c = 0$ . We have

$$\sum_{0 < s \leq t} [g(C(s)) - g(C(s-))] = \sum_{0 < s \leq t} [g(C(s-) + \Delta C(s)) - g(C(s-))]$$

or  $\sum_{0 < s \leq t} [g(C(s)) - g(C(s-))] = \int_{0+}^t \int_{\mathbb{R}} [g(C(s-) + y) - g(C(s-))] \mathcal{M}(ds, dy)$ .

▲

### 5.3 Itô's Formula for $\mathbb{R}^d$ -Valued Semimartingales

Itô's formulas are given for functions  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  of  $\mathbb{R}^d$ -valued semimartingales  $X$ . It is assumed that  $g$  has continuous second order partial derivatives.

**Theorem 5.3** *If the coordinates of  $X$  are continuous semimartingales, then  $g(X)$  is a continuous semimartingale and, for all  $t \geq 0$ , the integral and differential forms,*

$$g(X(t)) - g(X(0)) = \sum_{i=1}^d \int_0^t \frac{\partial g(X(s))}{\partial x_i} dX_i(s) + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 g(X(s))}{\partial x_i \partial x_j} d[X_i, X_j](s)$$

and

$$dg(X(t)) = \sum_{i=1}^d \frac{\partial g(X(t))}{\partial x_i} dX_i(t) + \frac{1}{2} \sum_{i,j}^d \frac{\partial^2 g(X(t))}{\partial x_i \partial x_j} d[X_i, X_j](t), \quad (5.15)$$

of Itô's formula hold with probability 1.

**Theorem 5.4** *If the coordinates of  $X$  are arbitrary semimartingales, then  $g(X)$  is a semimartingale and, for all  $t \geq 0$ , the integral form,*

$$\begin{aligned} g(X(t)) - g(X(0)) &= \sum_{i=1}^d \int_{0+}^t \frac{\partial g(X(s-))}{\partial x_i} dX_i(s) \\ &+ \frac{1}{2} \sum_{i,j=1}^d \int_{0+}^t \frac{\partial^2 g(X(s-))}{\partial x_i \partial x_j} d[X_i, X_j]^c(s) \\ &+ \sum_{0 < s \leq t} \left[ g(X(s)) - g(X(s-)) - \sum_{i=1}^d \frac{\partial g(X(s-))}{\partial x_i} \Delta X_i(s) \right]. \end{aligned} \quad (5.16)$$

of Itô's formula holds with probability 1.

The proof of these theorems follows from arguments similar to those employed to obtain Itô's formula for real-valued semimartingales ([15], Theorem 33, p. 74). The integral form of Itô's formula in (5.15) is a special case of (5.16) since  $X(s-) = X(s)$ ,  $\Delta X_i(s) = 0$ , and  $[X_i, X_j]^c(s) = [X_i, X_j](s)$  for continuous semimartingales.

*Example 5.6* Let  $X(t) = \exp[(c - \sigma^2/2)t + \sigma B(t)]$ ,  $t \geq 0$ , where  $B$  denotes a standard Brownian motion and  $c, \sigma \in \mathbb{R}$  are constants. This process is a continuous semimartingale satisfying the stochastic differential equation  $dX(t) = cX(t)dt + \sigma X(t)dB(t)$ ,  $t \geq 0$ , with initial state  $X(0) = 1$ . The process is referred to as geometric Brownian motion.  $\diamond$

*Proof* The differential form of Itô's formula in (5.15) applied to mapping  $(t, B(t)) \mapsto X(t) = g(t, B(t)) = \exp[(c - \sigma^2/2)t + \sigma B(t)]$  gives

$$dX(t) = \frac{\partial g(t, B(t))}{\partial t} dt + \frac{\partial g(t, B(t))}{\partial B(t)} dB(t) + \frac{1}{2} \frac{\partial^2 g(t, B(t))}{\partial B(t)^2} dt,$$

which yields the stated differential equation for  $X$  since  $\partial g/\partial t = (c - \sigma^2/2)X(t)$ ,  $\partial g/\partial B = \sigma X(t)$ , and  $\partial^2 g/\partial B^2 = \sigma^2 X(t)$ .  $\blacktriangle$

*Example 5.7* Let  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  be a real-valued function with continuous second order partial derivatives. Let  $B$  be an  $\mathbb{R}^d$ -valued Brownian motion whose coordinates  $B_i$  are independent Brownian motions starting at  $B_i(0) = x_i$ ,  $i = 1, \dots, d$ . The generator of  $B$  is

$$\mathcal{A}[g(x)] = \lim_{t \downarrow 0} \frac{E^x[g(B(t))] - g(x)}{t} = \frac{1}{2} \sum_{i=1}^d \frac{\partial^2 g(x)}{\partial x_i^2} = \frac{1}{2} \Delta g(x), \quad (5.17)$$

where  $E^x[\cdot] = E[\cdot | B(0) = x]$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ , and  $\Delta = \sum_{i=1}^d \partial^2/\partial x_i^2$  denotes the Laplace operator.  $\diamond$

*Proof* The Itô formula in (5.15) gives

$$g(B(t)) - g(B(0)) = \sum_{i=1}^d \int_0^t \frac{\partial g(B(s))}{\partial x_i} dB_i(s) + \frac{1}{2} \sum_{i=1}^d \int_0^t \frac{\partial^2 g(B(s))}{\partial x_i^2} ds$$

since  $d[B_i, B_j](s) = \delta_{ij} ds$ , where  $\delta_{ij} = 1$  for  $i = j$  and  $\delta_{ij} = 0$  for  $i \neq j$  (Example 4.21). The integrals in the first summation are  $\mathcal{F}_t$ -martingales starting at zero so that their expectation is zero. The integrals in the second summation can be defined as Riemann integrals and can be approximated by  $t \partial^2 g(B(\theta(\omega)t, \omega))/\partial x_i^2$ ,  $\theta(\omega) \in (0, 1)$ , for almost all  $\omega$ 's as  $t \downarrow 0$  so that their limits scaled by  $t$  converge to  $\partial^2 g(x)/\partial x_i^2$  as  $t \downarrow 0$ .  $\blacktriangle$

*Example 5.8* A stochastic process  $X$  is a standard Brownian motion if and only if it is a continuous local martingale with  $X(0) = 0$  and  $[X, X](t) = t$ . This fact is referred to as Lévy's characterization theorem ([12], Theorem 8.4.2).  $\diamond$

*Proof* If  $X$  is a Brownian motion, it has the stated properties. Suppose now that  $X$  is a continuous local martingale with  $X(0) = 0$  and  $[X, X](t) = t$  and set  $Z(t) = g(X(t), t) = \exp(iuX(t) + u^2 t/2)$ ,  $u \in \mathbb{R}$ . The Itô formula applied to  $Z(t) = g(X(t), t)$  gives

$$\begin{aligned}
Z(t) &= Z(0) + \int_0^t \left( \frac{\partial g(X(s), s)}{\partial x} dX(s) + \frac{\partial g(X(s), s)}{\partial s} ds \right) \\
&\quad + \frac{1}{2} \int_0^t \frac{\partial^2 g(X(s), s)}{\partial x^2} d[X, X](s) \\
&= 1 + iu \int_0^t Z(s) dX(s) + \frac{u^2}{2} \int_0^t Z(s) ds - \frac{u^2}{2} \int_0^t Z(s) d[X, X](s) \\
&= 1 + iu \int_0^t Z(s) dX(s).
\end{aligned}$$

Note that  $Z(t)$  is a complex-valued local martingale by a preservation property of the stochastic integral ([15], Theorem 17, p. 107). The martingale property,  $E[Z(t) | \mathcal{F}_s] = Z(s)$ ,  $t \geq s$ , implies  $E[e^{iu(X(t)-X(s))} | \mathcal{F}_s] = e^{-u^2(t-s)/2}$  for each  $u \in \mathbb{R}$  so that  $X(t) - X(s)$  is independent of  $\mathcal{F}_s$  and is normally distributed with mean zero and variance  $t - s$ . In summary,  $X$  starts at zero, is  $\mathcal{F}_t$ -adapted, has continuous samples, and has stationary Gaussian increments with mean zero and variance  $t - s$  that are independent of the past, so that it is a Brownian motion.  $\blacktriangle$

## 5.4 Itô and Stratonovich Integrals

The Itô and Stratonovich integrals with Brownian motion integrands and integrators have been defined in Examples 4.1 and 4.2. They are  $\int_0^t B(s) dB(s) = (B(t)^2 - t)/2$  and  $\int_0^t B(s) \circ dB(s) = B(t)^2/2$ , and are related by

$$\begin{aligned}
\int_0^t B(s) \circ dB(s) &= \int_0^t B(s) dB(s) + t/2 = \int_0^t B(s) dB(s) + \frac{1}{2} [B, B](t) \\
&= \int_0^t B(s) dB(s) + \frac{1}{2} [B, B]^c(t),
\end{aligned} \tag{5.18}$$

where the latter equality is valid since  $[B, B](t) = [B, B]^c(t)$ . In Example 4.4, we have noted that the Itô, the Stratonovich, and the path-by-path Riemann-Stieltjes definitions of  $\int N_- dN$  coincide. Note also that

$$\int_0^t N(s-) \circ dN(s) = \int_0^t N(s-) dN(s) = \int_0^t N(s-) dN(s) + \frac{1}{2} [N, N]^c(t) \tag{5.19}$$

holds since  $[N, N]^c(t) = 0$ .

**Definition 5.1** Let  $X$  and  $Y$  be semimartingales. The Stratonovich integral of  $X$  with respect to  $Y$  is

$$\int_0^t X(s-) \circ dY(s) = \int_0^t X(s-) dY(s) + \frac{1}{2} [X, Y]^c(t). \tag{5.20}$$

**Theorem 5.5** *If at least one of the semimartingales  $X$  and  $Y$  is continuous, then*

$$X(t)Y(t) - X(0)Y(0) = \int_{0+}^t X(s-) \circ dY(s) + \int_{0+}^t Y(s-) \circ dX(s). \quad (5.21)$$

*Proof* The integration by parts formula for semimartingales in (4.31) is

$$\int_{0+}^t X(s-) dY(s) = X(t)Y(t) - \int_{0+}^t Y(s-) dX(s) - [X, Y](t). \quad (5.22)$$

If  $X$  and/or  $Y$  is continuous, then  $[X, Y](t) = [X, Y]^c(t) + X(0)Y(0)$  and (5.22) becomes  $\int_{0+}^t X(s-) dY(s) = X(t)Y(t) - \int_{0+}^t Y(s-) dX(s) - [X, Y]^c(t) - X(0)Y(0)$ . The formula in (5.21) results by adding  $[X, Y]^c(t)/2$  to both sides of the latter equation and using Definition 5.1.  $\blacktriangle$

**Theorem 5.6** *If  $X$  is an  $\mathbb{R}^d$ -valued semimartingales and  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  has continuous second order partial derivatives, then  $g(X)$  is a semimartingale and the formula*

$$\begin{aligned} g(X(t)) - g(X(0)) &= \sum_{i=1}^d \int_{0+}^t \frac{\partial g}{\partial x_i}(X(s-)) \circ dX_i(s) \\ &+ \sum_{0 < s \leq t} \left[ g(X(s)) - g(X(s-)) - \sum_{i=1}^d \frac{\partial g(X(s-))}{\partial x_i} \Delta X_i(s) \right]. \end{aligned} \quad (5.23)$$

holds ([15], Theorem 21, p. 222).

The formulas in Theorems 5.4 and 5.6 give the relationship

$$\begin{aligned} \sum_{i=1}^d \int_{0+}^t \frac{\partial g}{\partial x_i}(X(s-)) \circ dX_i(s) &= \sum_{i=1}^d \int_{0+}^t \frac{\partial g(X(s-))}{\partial x_i} dX_i(s) \\ &+ \frac{1}{2} \sum_{i,j=1}^d \int_{0+}^t \frac{\partial^2 g(X(s-))}{\partial x_i \partial x_j} d[X_i, X_j]^c(s) \end{aligned} \quad (5.24)$$

between the Stratonovich and Itô integrals. Generally, the state equations for physical systems are driven by colored (non-white) noise so that they are interpreted as Stratonovich stochastic differential equations. Relationships as in (5.20) and (5.24) can be used to transform these equations into Itô stochastic differential equations that may be simpler to solve. We will return to this topic in Sect. 5.5.1.1 (Theorem 5.10).

## 5.5 Applications

We have already seen some applications of Itô's formula. In Examples 5.1 and 5.2, the formula has been employed to calculate the stochastic integrals  $\int B dB$  and  $\int N_- dN$ . Stratonovich's integral has been defined and related to Itô's integral in Sect. 5.4. The



differential form of Itô's formula has been used in Example 5.6 to construct a stochastic differential equation for a real-valued process, referred to as geometric Brownian motion. The infinitesimal generator of Brownian motion needed to construct local solutions for the Laplace equation has been derived in Example 5.7.

This section provides additional applications of Itô's formula. They include essentials on stochastic differential equations driven by Brownian motion and semimartingales, Tanaka's formula that is used to solve locally partial differential equations with mixed boundary conditions, and Girsanov's theorem providing a framework for developing efficient Monte Carlo algorithms.

### 5.5.1 Stochastic Differential Equations

It was shown in Example 5.6 that  $X(t) = \exp[(c - \sigma^2/2)t + \sigma B(t)]$  satisfies the differential equation  $dX(t) = cX(t)dt + \sigma X(t)dB(t)$  with initial state  $X(0) = 1$ . The meaning of this equation is given by its integral form  $X(t) = X(0) + c \int_0^t X(s)ds + \sigma \int_0^t X(s)dB(s)$ , where  $\int_0^t X(s)ds$  and  $\int_0^t X(s)dB(s)$  are Riemann and Itô integrals. The differential and integral representations of  $X(t)$  are called stochastic differential and integral equations, respectively.

This section considers stochastic differential equations of the type

$$dX(t) = a(X(t-), t)dt + b(X(t-), t)dY(t), \quad t \geq 0, \quad (5.25)$$

with integral form

$$X(t) = X(0) + \int_0^t a(X(s-), s)ds + \int_0^t b(X(s-), s)dY(s), \quad t \geq 0, \quad (5.26)$$

where  $a, b$  are  $(d, 1)$ ,  $(d, d')$ -matrices whose entries are real-valued Borel measurable functions,  $Y$  is an  $\mathbb{R}^{d'}$ -valued semimartingale,  $X$  is an  $\mathbb{R}^d$ -valued stochastic process. The first and second integrals in (5.26) are Riemann-Stieltjes and stochastic or Itô integrals. As in Sect. 3.7.6.4, we refer to the formal derivatives of Brownian motion, compound Poisson, Lévy, and semimartingale processes as Gaussian, Poisson, Lévy, and semimartingale white noise processes.

#### 5.5.1.1 Gaussian White Noise

Let  $Y$  in (5.26) be an  $\mathbb{R}^{d'}$ -valued Brownian motion process  $B = (B_1, \dots, B_{d'})$ , where  $B_i$ ,  $i = 1, \dots, d'$ , are independent real-valued Brownian motions, so that  $X$  satisfies the stochastic integral equation

$$X(t) = X(0) + \int_0^t a(X(s), s)ds + \int_0^t b(X(s), s)dB(s), \quad t \geq 0. \quad (5.27)$$

The process  $X$ , the matrix  $a$ , and the matrix  $b$  or  $b b'$  in (5.27) are called diffusion process, drift or drift coefficient, and diffusion or diffusion coefficient, respectively.  $X(s-)$  in (5.26) is changed into  $X(s)$  since diffusion processes have continuous samples, as it will see later in this section.

The solution  $X$  of (5.27) is a Markov process since  $B$  has independent increments so that, for given  $X(t_0) = x$ ,  $t_0 \geq 0$ , future states

$$X(t) = x + \int_{t_0}^t a(X(s), s) ds + \int_{t_0}^t b(X(s), s) dB(s), \quad t \geq t_0,$$

are independent of past states  $X(u)$ ,  $u < t_0$ . The converse is not true. For example, the compound Poisson process is Markov but is not a diffusion process. Moreover,  $X$  has the strong Markov property, that is, if  $T$  is an  $\mathcal{F}_t = \sigma(X(0), B(s) : 0 \leq s \leq t)$ -stopping time, the process  $X(T + \tau)$ ,  $\tau \geq 0$ , depends only on  $X(T)$ .

The solution of (5.27) can be defined in the strong and the weak sense. If the initial state  $X(0)$  is deterministic, a strong solution is a stochastic process  $X(t)$ ,  $t \geq 0$ , such that (1)  $X$  is adapted to the filtration  $\mathcal{F}_t = \sigma(B(s), 0 \leq s \leq t)$  generated by  $B$ , (2)  $X$  is a function of the samples of  $B$  and of the coefficients  $a$  and  $b$ , and (3) the Riemann-Stieltjes and Itô integrals in (5.27) are well defined at all times ([13], p. 137). If  $X(0)$  is random,  $\mathcal{F}_t$  needs to be extended to  $\sigma(X(0), \mathcal{F}_t)$ . We construct approximate strong solutions in Monte Carlo studies since the noise version needs to be specified to generate samples of  $B$  and samples of  $X$  are calculated from samples of  $B$  by integrating (5.27) numerically.

A weak solution of (5.27) is a pair of adapted processes  $(\tilde{B}, \tilde{X})$  defined on a filtered probability space  $(\Omega, \mathcal{H}, (\mathcal{H}_t)_{t \geq 0}, P)$  such that  $\tilde{B}$  is a version of  $B$  and the pair  $(\tilde{B}, \tilde{X})$  satisfies (5.27). The weak solution is completely defined by the initial conditions, the functions  $a$  and  $b$ , and the finite dimensional distributions of  $B$ . The particular version of the input does not have to be specified. Note that a strong solution is a weak solution but the converse is not generally true.

Uniqueness for the solution of (5.27) can be defined in two ways. A strong solution is said to be unique in the strong sense if two different solutions of this equation have the same samples except on a subset of  $\Omega$  of measure zero, that is, they are indistinguishable processes. Two solutions, weak or strong, of (5.27) are unique in the weak sense if they have the same finite dimensional distributions, that is, they are versions ([3], Sect. 10.4, [13], Sect. 3.2.1, [14], Sect. 5.3).

*Example 5.9* Let  $X(t)$  be a real-valued process defined by the stochastic differential equation

$$dX(t) = \text{sign}(X(t)) dB(t), \quad t \geq 0,$$

where  $X(0) = 0$ ,  $B$  is a Brownian motion starting at zero, and  $\text{sign}(x) = -1, 0$ , and  $1$  for  $x < 0$ ,  $x = 0$ , and  $x > 0$ , respectively. This equation has a weak solution ([3], Sect. 7.3) that is unique in the weak sense ([3], pp. 248–249) but it is not unique in the strong sense.  $\diamond$

*Proof* Let  $X(\cdot, \omega)$  be a sample of  $X(t)$  corresponding to a sample  $B(\cdot, \omega)$  of  $B$ , that is,  $X(t, \omega) - X(0, \omega) = \int_0^t \text{sign}(X(s, \omega)) dB(s, \omega)$ . Then  $\tilde{X} = -X$  satisfies the equation  $\tilde{X}(t, \omega) - \tilde{X}(0, \omega) = \int_0^t \text{sign}(\tilde{X}(s, \omega)) dB(s, \omega)$ . If  $X(0) = 0$ , then  $X(\cdot, \omega)$  and  $\tilde{X}(t, \omega)$  are solutions corresponding to the same sample  $B(\cdot, \omega)$  of  $B$ . The processes  $X$  and  $\tilde{X}$  have the same probability law but their samples differ.  $\blacktriangle$

*Example 5.10* Let  $X$  be the solution of  $dX(t) = -\alpha X(t) dt + \beta dB(t)$ ,  $t \geq 0$ , where  $\alpha > 0$  and  $\beta$  are some constants,  $X(0) \sim N(\mu(0), \gamma(0))$ ,  $B$  is a Brownian motion, and  $X(0)$  is independent of  $B$ . This real-valued diffusion process, called the Ornstein–Uhlenbeck process, is a special case of (5.27) with  $d = d' = 1$ , drift  $a(x) = -\alpha x$ , and diffusion  $b(x)^2 = \beta^2$ . The theorems in the following section guarantee the existence and the uniqueness of the solution  $X$ .

Samples of the strong solution of  $dX(t) = -\alpha X(t) dt + \beta dB(t)$  can be calculated from samples  $B(\cdot, \omega)$  of  $B$  and the recurrence formula

$$X(t + \Delta t, \omega) = X(t, \omega) e^{-\alpha \Delta t} + \beta \int_t^{t+\Delta t} e^{-\alpha(t+\Delta t-s)} dB(s, \omega)$$

giving  $X$  at the end of  $[t, t + \Delta t]$  from its value at the beginning of this time interval and the input in  $[t, t + \Delta t]$ . The weak solution of this equation is given by a pair of a Brownian motion  $\tilde{B}$  and a Gaussian process  $\tilde{X}$  with mean  $\mu(t) = E[X(t)] = \mu(0) e^{-\alpha t}$ , variance  $\gamma(t) = E[\tilde{X}(t)^2] = \gamma(0) e^{-2\alpha t} + (1 - e^{-2\alpha t}) \beta^2 / (2\alpha)$ , and covariance  $c(t, s) = E[\tilde{X}(t) \tilde{X}(s)] = \gamma(s \wedge t) e^{-\alpha|t-s|}$ .  $\diamond$

*Proof* The formula relating  $X(t + \Delta t)$  to  $X(t)$  shows that the Ornstein–Uhlenbeck process is Gaussian, so that the second moment properties of  $X$  define its finite dimensional distributions. The mean equation can be obtained by averaging the defining equation for  $X$ . The stochastic integral equation

$$X(t) = X(s) - \alpha \int_s^t X(u) du + \beta \int_s^t dB(u), \quad t \geq s,$$

multiplied by  $X(s)$  and averaged gives

$$E[X(t)X(s)] = E[X(s)^2] - \alpha E\left[\int_s^t X(u)X(s) du\right] + \beta E\left[\int_s^t X(s) dB(u)\right]$$

or  $r(t, s) = r(s, s) - \alpha \int_s^t r(u, s) du$  by using the Fubini theorem, the independence of  $X(s)$  from future increments of the Brownian motion, and  $E[ dB(u)] = 0$ , where  $r(t, s) = E[X(t)X(s)]$ . This equation gives  $\partial r(t, s) / \partial t = -\alpha r(t, s)$  by differentiation with respect to  $t$  so that  $r(t, s) = r(s, s) e^{-\alpha(t-s)}$ .  $\blacktriangle$

We state without proof two theorems giving conditions for the existence and uniqueness of solutions of stochastic differential equations. The statements of these theorems involve concepts that are clarified by the following definitions.

**Definition 5.2** Let  $a(x)$  and  $b(x)$  be  $(d, 1)$  and  $(d, d')$  matrices whose entries are functions of  $x \in \mathbb{R}^d$ . These matrices satisfy the uniform Lipschitz conditions if there exists a constant  $c > 0$  such that

$$\|a(x_1) - a(x_2)\| \leq c \|x_1 - x_2\| \quad \text{and} \quad \|b(x_1) - b(x_2)\|_m \leq c \|x_1 - x_2\|, \quad (5.28)$$

where  $\|\xi\| = (\sum_{i=1}^d \xi_i^2)^{1/2}$  is the Euclidean norm,  $\|b\|_m = (\sum_{i=1}^d \sum_{j=1}^{d'} b_{ij}^2)^{1/2}$  denotes a matrix norm, and  $x_1, x_2 \in \mathbb{R}^d$ .

**Definition 5.3** The matrices  $a$  and  $b$  are said to be locally Lipschitz if, for each  $\alpha > 0$ , there is a constant  $c_\alpha > 0$  such that

$$\|a(x_1) - a(x_2)\| \leq c_\alpha \|x_1 - x_2\| \quad \text{and} \quad \|b(x_1) - b(x_2)\|_m \leq c_\alpha \|x_1 - x_2\| \quad (5.29)$$

for  $x_1, x_2 \in \mathbb{R}^d$  satisfying the condition  $\|x_1\|, \|x_2\| < \alpha$ .

For example, the function  $f(x) = \sqrt{x^2 + 5}$ ,  $x \in \mathbb{R}$ , is uniform Lipschitz since it is differentiable, the absolute value of its derivative  $f'(x) = x/\sqrt{x^2 + 5}$  is smaller than 1, and there is at least a point  $\xi$  in any interval  $(x_1, x_2)$  such that  $f'(\xi) = (f(x_2) - f(x_1))/(x_2 - x_1)$  by the mean value theorem ([2], Theorem 20.3). Also,  $f(x) = |x|$ ,  $x \in \mathbb{R}$ , is uniform Lipschitz since  $||x_1| - |x_2|| \leq |x_1 - x_2|$  by the reverse triangle inequality. The function  $f(x) = x^2$ ,  $x \in \mathbb{R}$ , is not uniform Lipschitz since its rate of change increases indefinitely as  $|x| \rightarrow \infty$ , but it is locally Lipschitz. The function  $f(x) = x^{3/2} \sin(1/x)1(x \neq 0)$ ,  $x \in [0, 1]$ , is an example of a differentiable function on a compact that is not locally Lipschitz since its derivative is not bounded.

**Definition 5.4** The matrices  $a$  and  $b$  satisfy the growth condition if there exists a constant  $k > 0$  such that

$$x \cdot a(x) \leq k(1 + \|x\|^2) \quad \text{and} \quad \|b(x)\|_m^2 \leq k(1 + \|x\|^2), \quad (5.30)$$

where  $x \cdot a(x) = \sum_{i=1}^d x_i a_i(x)$ . The condition  $x \cdot a(x) \leq k(1 + \|x\|^2)$  is weaker than the typically stated growth condition,  $\|a(x)\|^2 \leq k'(1 + \|x\|^2)$ , because  $(x - a(x)) \cdot (x - a(x)) \geq 0$  so that  $x \cdot a(x) \leq (\|x\|^2 + \|a(x)\|^2)/2 \leq (k' + 1)(1 + \|x\|^2)/2$ .

The existence and uniqueness theorems for the solution of (5.27) are stated for the case in which the drift and diffusion coefficients do not depend explicitly on time. This is not restrictive since, if the drift and/or diffusion coefficients of an equation depend on time, we can apply these theorems to the state vector  $\tilde{X}(t) = (\tilde{X}^{(1)}(t) = X(t), \tilde{X}^{(2)}(t) = t) \in \mathbb{R}^{d+1}$  with  $\tilde{X}^{(1)} = X$  defined by (5.27) and  $d\tilde{X}^{(2)}(t) = dt$  with the initial condition  $\tilde{X}^{(2)}(0) = 0$ . Theorems dealing directly with drift and diffusion coefficients that depend explicitly on time are available ([10], Sect. 4.5, [11], Theorem 7.1.1, p. 195, [14], Theorem 5.2.1, p. 66, [16], Sect. 4.2).

**Theorem 5.7** *If the drift and diffusion coefficients in (5.27) defined on a time interval  $[0, \tau]$  do not depend on time explicitly, are bounded functions and satisfy the uniform Lipschitz conditions in (5.28),  $B$  is a Brownian motion martingale on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ ,  $B(0) = 0$ , and  $X(0)$  is  $\mathcal{F}_0$ -measurable, then*

(1) there exists a strong solution  $X$  for (5.27) that is unique in the strong sense, (2)  $X$  is a  $\mathcal{B}([0, \infty)) \times \mathcal{F}$ -measurable,  $\mathcal{F}_t$ -adapted, and continuous process such that  $\sup_{0 \leq t \leq \tau} E[X(t)^2] < \infty$ , and (3) the law of  $X$  is uniquely determined by the drift and diffusion coefficients and the laws of  $B$  and  $X(0)$  ([3], Theorem 10.5).

**Theorem 5.8** *The statements in Theorem 5.7 also hold if the assumptions that the drift and diffusion coefficients in (5.27) are bounded, uniform Lipschitz functions are replaced with the assumptions that the drift and diffusion coefficients are locally Lipschitz functions satisfying the growth conditions in (5.30) ([3], Theorem 10.6, [18], Theorem 9.1, [14], Theorem 5.2.1).*

*Example 5.11* The stochastic differential equation

$$dX(t) = cX(t)dt + \sigma X(t)dB(t), \quad t \in [0, \tau], \quad (5.31)$$

has the unique strong solution

$$X(t) = X(0) \exp[(c - \sigma^2/2)t + \sigma B(t)], \quad (5.32)$$

called the geometric Brownian motion process (Example 5.6).  $\diamond$

*Proof* That the above stochastic differential equation has a unique strong solution follows from Theorem 5.8 since  $a(x) = cx$  and  $b(x) = \sigma x$  are locally Lipschitz and satisfy the growth conditions.

Theorem 5.7 cannot be applied directly since  $a(x)$  and  $b(x)$  are not bounded, but the original problem can be modified to satisfy the requirements of this theorem. Let  $X$  be the solution of a stochastic differential equation  $dX(t) = a(X(t))dt + b(X(t))dB(t)$ , where the coefficients  $a, b$  satisfy the uniform Lipschitz conditions but may not be bounded. For  $\xi > 0$  define the function  $[x]_\xi = -1(x < -\xi)\xi + 1(-\xi \leq x \leq \xi)x + 1(x > \xi)\xi$  and consider the stochastic differential equation

$$d\tilde{X}(t) = a([X(t)]_\xi)dt + b([X(t)]_\xi)dB(t).$$

Since the functions  $a([\cdot]_\xi)$  and  $b([\cdot]_\xi)$  satisfy the conditions of Theorem 5.7,  $\tilde{X}$  exists and is unique in the strong sense for each  $\xi > 0$ . Since  $\xi > 0$  is arbitrary, (5.31) has a unique solution in the strong sense.  $\blacktriangle$

**Theorem 5.9** *If the conditions of Theorem 5.7 are satisfied, the solution  $X$  of (5.27) is a semimartingale with the representation*

$$X(t) = X(0) + A(t) + M(t), \quad (5.33)$$

where  $A(t) = \int_0^t a(X(s))ds$  is an adapted process with samples of finite variation on compacts,  $M(t) = \int_0^t b(X(s))dB(s)$  is an  $\mathcal{F}_t$ -square integrable martingale,  $A(0) = 0$ , and  $M(0) = 0$  ([3], Theorem 10.5, p. 228).

*Example 5.12* If the conditions of Theorem 5.7 are satisfied and  $d = d' = 1$  in (5.27), then this equation has a unique strong solution and

$$g(X(t)) - g(X(0)) = \int_0^t g'(X(s)) dX(s) + \frac{1}{2} \int_0^t g''(X(s)) b(X(s))^2 ds$$

for any function  $g \in C^2(\mathbb{R})$ .  $\diamond$

*Proof* Since  $X$  is a continuous semimartingale, Itô's formula in (5.3) can be applied to  $g(X)$ . This formula involves the quadratic variation process  $[X, X]$ , where  $X(t) = Z + A(t) + M(t)$ ,  $Z = X(0)$ ,  $A(t) = \int_0^t a(X(s)) ds$ , and  $M(t) = \int_0^t b(X(s)) dB(s)$ . We have  $[X, X] = [Z, Z] + [Z, A] + [Z, M] + [A, Z] + [A, A] + [A, M] + [M, Z] + [M, A] + [M, M]$  by the linearity of quadratic covariation. All quadratic covariations with arguments  $Z$  and  $A$  or  $M$  are zero since  $Z$  is a constant process and  $A(0) = M(0) = 0$ . The quadratic variation of  $A$  and the quadratic covariation of  $A$  and  $M$  are also zero so that

$$[X, X] = Z^2 + [M, M] = Z^2 + \int_0^t b(X(s))^2 d[B, B](s) = Z^2 + \int_0^t b(X(s))^2 ds$$

implying  $d[X, X](t) = b(X(t))^2 dt$  since  $Z^2$  is a constant process.  $\blacktriangle$

*Example 5.13* Let  $X$  be the diffusion process in Example 5.12 and  $g : \mathbb{R} \rightarrow \mathbb{R}$  be an increasing function with continuous second order derivative. The memoryless transformation  $Y(t) = g(X(t))$  of  $X$  is a diffusion process defined by

$$dY(t) = a_Y(Y(t)) dt + b_Y(Y(t)) dB(t) \quad \text{where}$$

$$a_Y(y) = g'(x) a(x) + \frac{1}{2} g''(x) b(x)^2, \quad b_Y(y) = g'(x) b(x), \quad \text{and } x = g^{-1}(y).$$

For example, let  $X$  be a real-valued diffusion process with drift  $a(x) = -x$  and diffusion  $b(x) = 1$ . The stochastic differential equation for  $Y = g(X) = X^3$  is

$$dY(t) = 3(-Y(t) + |Y(t)|^{1/3} \text{sign}(Y(t))) dt + 3|Y(t)|^{2/3} dB(t)$$

since  $x = g^{-1}(y) = |y|^{1/3} \text{sign}(y)$ . Sample paths of  $Y$  can be obtained from samples of  $X$  and the memoryless transformation  $Y = X^3$  or can be generated directly from the stochastic differential equation for  $Y$ .

Memoryless transformations can be used to generate diffusion processes with specified marginal distributions. For example,  $Y(t) = F^{-1}(\Phi(X(t)))$  is a diffusion process with marginal distribution  $F$ , where  $\Phi$  denotes the distribution of the standard Gaussian variable and  $X(t)$  can be a stationary Ornstein–Uhlenbeck process with mean 0 and variance 1. The probability law of  $Y(t)$  is defined by the properties of  $X(t)$  and the mapping  $X(t) \mapsto Y(t) = F^{-1}(\Phi(X(t)))$ . Note that  $Y(t)$  is a translation process of the type discussed in Sect. 3.7.2.  $\diamond$

*Proof* The differential form of Itô's formula in (5.3) gives

$$dY(t) = \left[ g'(X(t)) a(X(t)) + \frac{1}{2} g''(X(t)) b(X(t))^2 \right] dt + g'(X(t)) b(X(t)) dB(t).$$

Since  $x = g^{-1}(y)$  exists,  $Y$  is a diffusion process satisfying the stochastic differential equation

$$dY(t) = \left[ g'(g^{-1}(Y(t))) a(g^{-1}(Y(t))) + \frac{1}{2} g''(g^{-1}(Y(t))) b(g^{-1}(Y(t)))^2 \right] dt + g'(g^{-1}(Y(t))) b(g^{-1}(Y(t))) dB(t)$$

with the stated drift and diffusion coefficients. ▲

*Example 5.14* Let  $X$  be defined by  $dX(t) = -\alpha X(t) dt + \beta dB(t)$ , where  $B$  is a Brownian motion and  $\alpha > 0$  and  $\beta$  are constants. The moments  $\mu(q; t) = E[X(t)^q]$  of  $X(t)$  satisfy the ordinary differential equation

$$\dot{\mu}(q; t) = -\alpha q \mu(q; t) + \frac{\beta^2 q(q-1)}{2} \mu(q-2; t), \quad q = 1, 2, \dots, \quad (5.34)$$

where  $\dot{\mu}(q; t) = d\mu(q; t)/dt$  and  $\mu(q; t) = 0$  for  $q \leq -1$  by convention. Note that  $X(t)$  has moments of any order since it is a Gaussian process. ◇

*Proof* The drift and diffusion coefficients of  $X$  are locally Lipschitz and satisfy the growth conditions so that there exists a unique, adapted, and continuous solution if  $X(0) \in \mathcal{F}_0$ . The integral form of Itô's formula is

$$g(X(t)) - g(X(0)) = \int_0^t g'(X(s)) dX(s) + \frac{\beta^2}{2} \int_0^t g''(X(s)) ds$$

for  $g \in C^2(\mathbb{R})$ . The expectation of this equation gives

$$E[g(X(t))] - E[g(X(0))] = -\alpha \int_0^t E[g'(X(s)) X(s)] ds + \frac{\beta^2}{2} \int_0^t E[g''(X(s))] ds$$

since  $\int_0^t g'(X(s)) dB(s)$  is a martingale starting at zero so that its expectation is zero and the solution  $X$  is a  $\mathcal{B}[0, \infty) \times \mathcal{F}$ -measurable function and  $P$ -integrable so that  $E[\int_0^t g'(X(s)) X(s) ds] = \int_0^t E[g'(X(s)) X(s)] ds$  and  $E[\int_0^t g''(X(s)) ds] = \int_0^t E[g''(X(s))] ds$  by Fubini's theorem. For  $g(x) = x^q$ , we have

$$\mu(q; t) - \mu(q; 0) = -\alpha q \int_0^t \mu(q; s) ds + \frac{\beta^2 q(q-1)}{2} \int_0^t \mu(q-2; s) ds$$

which gives (5.34) by differentiation with respect to  $t$ . ▲

*Example 5.15* Let  $X$  be the Ornstein–Uhlenbeck process in Example 5.14. The characteristic function  $\varphi(u; t) = E[\exp(iuX(t))]$ ,  $u \in \mathbb{R}$ , of  $X(t)$  satisfies the partial differential equation

$$\frac{\partial \varphi}{\partial t} = -\alpha u \frac{\partial \varphi}{\partial u} - \frac{\beta^2 u^2}{2} \varphi. \quad (5.35)$$

The stationary solution  $\varphi_s(u) = \lim_{t \rightarrow \infty} \varphi(u; t) = \exp(-\beta^2 u^2/(4\alpha))$  of this equation satisfies the ordinary differential equation  $\alpha u \varphi'_s(u) + \beta^2 u^2 \varphi_s(u)/2 = 0$  with solution  $\varphi_s(u) = \exp(-\beta^2 u^2/(4\alpha))$ , so that  $X(t) \sim N(0, \beta^2/(2\alpha))$  as  $t \rightarrow \infty$ .  $\diamond$

*Proof* Itô's formula applied to the real and imaginary parts of  $g(X) = \exp(iuX(t))$ ,  $u \in \mathbb{R}$ , gives

$$e^{iuX(t)} - e^{iuX(0)} = \int_0^t (iu) e^{iuX(s)} dX(s) + \frac{\beta^2}{2} \int_0^t (iu)^2 e^{iuX(s)} ds,$$

by adding these contributions. The expectation of this equation is

$$\varphi(u; t) - \varphi(u; 0) = -iu\alpha \int_0^t E[e^{iuX(s)} X(s)] ds - \frac{\beta^2 u^2}{2} \int_0^t E[e^{iuX(s)}] ds$$

by using arguments as in the previous example. The differential equation for the characteristic function results by differentiating the above formula with respect to  $t$  and the equality  $\partial^k \varphi(u; t)/\partial u^k = E[(iX(t))^k \exp(iuX(t))]$ .

Since  $\varphi_s$  is the solution of the ordinary differential equation  $\varphi'_s(u)/\varphi_s(u) = -\beta^2 u/(2\alpha)$ , it is  $\ln(\varphi_s(u)) = -\beta^2 u^2/(4\alpha) + c$  or  $\varphi_s(u) = c' \exp(-\beta^2 u^2/(4\alpha))$ , where  $c$  and  $c'$  are constants. The condition  $\varphi_s(0) = 1$  implies  $c' = 1$ , so that the marginal distribution of  $X(t)$  approaches that of  $N(0, \beta^2 u^2/(2\alpha))$  as  $t \rightarrow \infty$ .  $\blacktriangle$

*Example 5.16* Let  $X$  be defined by the stochastic differential equation  $dX(t) = -\alpha X(t) dt + \sigma X(t) dB(t)$ ,  $t \geq 0$ , where  $\alpha, \sigma$  are constants and  $B$  denotes a Brownian motion. Then

$$\mu(q; t) = E[X(t)^q] = \mu(q; 0) \exp\left[\left(-\alpha + \frac{(q-1)\sigma^2}{2}\right)qt\right]$$

for any integer  $q \geq 0$ . The moment of order  $q$  of  $X$  approaches zero and  $\pm\infty$  as  $t \rightarrow \infty$  if  $\alpha > \alpha^*(q)$  and  $\alpha < \alpha^*(q)$ , respectively, where  $\alpha^*(q) = (q-1)\sigma^2/2$ . If  $\alpha = \alpha^*(q)$ , then  $\mu(q; t) = \mu(q; 0)$  is time invariant.  $\diamond$

*Proof* The moment equation is  $\dot{\mu}(q; t) = -\alpha q \mu(q; t) + (1/2) q(q-1) \sigma^2 \mu(q; t)$ , and has the stated solution. If  $-\alpha + (q-1)\sigma^2/2 < 0$ , or equivalently,  $\alpha > \alpha^*(q)$ , then  $\mu(q; t)$  decreases to zero as  $t \rightarrow \infty$ . If  $\alpha < \alpha^*(q)$ ,  $\mu(q; t)$  converges to  $\pm\infty$  as  $t \rightarrow \infty$  depending on the sign of  $\mu(q; 0)$ .

We have seen in Examples 5.6 and 5.11 that the solution of  $dX(t) = -\alpha X(t) dt + \sigma X(t) dB(t)$  is the process  $X(t) = X(0) \exp[(-(\alpha + \sigma^2/2)t + \sigma B(t))]$ , referred to as geometric Brownian motion. If  $X(0)$  is independent of  $B$ , then  $E[X(t)^q] = E[X(0)^q] E[\exp(q(-(\alpha + \sigma^2/2)t + \sigma B(t)))]$ , which coincides with  $\mu(q; t)$  since  $E[\exp(q\sigma B(t))] = \exp(q^2 \sigma^2 t/2)$ .  $\blacktriangle$

*Example 5.17* Let  $X$  be the diffusion process in Example 5.12 and  $h: \mathbb{R} \rightarrow \mathbb{R}$  be a function with continuous first order derivative. Then,

$$\int_0^t h(X(s)) \circ dX(s) = \int_0^t h(X(s)) dX(s) + \frac{1}{2} \int_0^t h'(X(s)) b(X(s))^2 ds,$$



where  $\int_0^t h(X(s)) \circ dX(s)$  and  $\int_0^t h(X(s)) dX(s)$  denote the Stratonovich and Itô integrals, respectively. This relationship follows from (5.20) and Theorem 4.13, which give  $d[h(X), X]^c(t) = h'(X) d[X, X](t) = h'(X(t)) b(X(t))^2 dt$ .  $\diamond$

*Example 5.18* Let  $X$  be the solution of the Itô integral equation

$$X(t) = X(0) + \int_0^t a(X(s), s) ds + \int_0^t b(X(s), s) dB(s), \quad (5.36)$$

where  $b(x, t)$  has continuous second and first order partial derivatives with respect to  $x$  and  $t$ , respectively. The process also satisfies the Stratonovich integral equation

$$X(t) = X(0) + \int_0^t \tilde{a}(X(s), s) ds + \int_0^t b(X(s), s) \circ dB(s), \quad (5.37)$$

where  $\tilde{a}(x, t) = a(x, t) - (1/2) b(x, t) [\partial b(x, t)/\partial x]$ .  $\diamond$

*Proof* The Itô formula applied to  $U(t) = b(X(t), t)$  gives

$$dU(t) = \frac{\partial U}{\partial x} (a(X(t), t) dt + b(X(t), t) dB(t)) + \frac{\partial U}{\partial t} dt + \frac{1}{2} \frac{\partial^2 U}{\partial x^2} d[X, X](t)$$

so that

$$d[U, B](t) = \left[ \frac{\partial U}{\partial x} b(X(t), t) dB(t), dB(t) \right] = b(X(t), t) \frac{\partial U}{\partial x} dt.$$

Since  $[U, B] = [U, B]^c$ , the relationship between the Itô and Stratonovich integrals in (5.20) becomes

$$\int_0^t U(s) \circ dB(s) = \int_0^t U(s) dB(s) + \frac{1}{2} \int_0^t b(X(s), s) \frac{\partial U}{\partial x} ds,$$

or, equivalently,

$$\int_0^t b(X(s), s) \circ dB(s) = \int_0^t b(X(s), s) dB(s) + \frac{1}{2} \int_0^t b(X(s), s) \frac{\partial b(X(s), s)}{\partial x} ds,$$

so that (5.36) becomes

$$X(t) = X(0) + \int_0^t \left[ a(X(s), s) - \frac{1}{2} \frac{\partial b(X(s), s)}{\partial x} \right] ds + \int_0^t b(X(s), s) \circ dB(s),$$

by replacing  $\int_0^t b(X(s), s) dB(s)$  with its expression in the previous formula. The result in (5.37) is referred to as the Stratonovich differential equation.  $\blacktriangle$

The stochastic differential equations considered in this chapter are driven by white noise processes that have infinite variance and can be interpreted as, for example, formal derivatives of Brownian motion processes. White noise is a mathematical

abstraction that has a constant spectral density over the entire real line, according to our previous formal calculations. In contrast, the differential equations describing the behavior of physical systems are driven by processes that may have a flat spectral density over a broad frequency band but their variance is finite. These processes cannot be viewed as the formal derivatives of Brownian motions, and are called colored noises to distinguish them from white noises. The significant difference between white and colored noises implies that the theory of stochastic differential equations cannot be used directly to solve practical problems. The theory can be used in applications if the drift of a differential equation with colored noise is modified and the colored noise is replaced with white noise. The Wong-Zakai theorem provides the mapping from equations with colored noise to stochastic differential equations.

A two-step algorithm can be used to solve practical problems. First, a differential equation driven by colored noise defining the evolution of the state  $X(t)$  of a physical system is developed based on physics, data, and any other available information. Second, the resulting equation is mapped into a stochastic differential equation with white noise such the two equations have the same solution  $X(t)$ . It is preferable to deal with differential equations driven by white rather than colored noise since they can be solved efficiently by using the tools of stochastic calculus.

The subsequent example shows that, generally, it is not possible to construct stochastic differential equations with white noise whose solutions are similar to those of differential equations with colored noise by using heuristic arguments. Following the example, we state the Wong-Zakai theorem, which provides a mapping between Itô and Stratonovich equations that share the same solutions, that is, between differential equations with white and colored noise inputs.

*Example 5.19* Let  $Y(t)$  be real-valued process satisfying the differential equation

$$\dot{Y}(t) = c Y(t) + \sigma Y(t) V(t), \quad t \geq 0,$$

where  $Y(0) = x \neq 0$ ,  $c$  and  $\sigma$  are constants, and  $V(t)$  is a colored Gaussian noise with continuous samples. The solution of this equation is

$$Y(t) = x \exp \left[ c t + \sigma \int_0^t V(s) ds \right], \quad t \geq 0,$$

by classical calculus, where  $\int_0^t V(s) ds$  is a path-by-path Riemann integral.

Suppose  $V$  is defined by  $dV(t) = -\alpha V(t) dt + \sqrt{2\alpha} dB(t)$ , where  $\alpha > 0$  and  $V(0) \sim N(0, 1)$  is independent of the Brownian motion  $B$ . Then  $V(t)$  is stationary Gaussian process with mean 0, covariance function  $E[V(t + \tau)V(t)] = \exp(-\alpha|\tau|)$ , spectral density  $s(\nu) = \alpha / [\pi(\alpha^2 + \nu^2)]$ , and a.s. continuous samples. Since for large values of  $\alpha$  the covariance of  $V(t)$  is nearly a  $\delta$ -function, it may be tempting to write

$$\tilde{Y}(t) = \lim_{\alpha \rightarrow \infty} Y(t) \simeq x \exp \left[ c t + \sigma \int_0^t dB(s) \right] = x \exp[c t + \sigma B(t)]$$

by using the approximation  $V(s) ds \simeq dB(s)$ . This may suggest that  $\tilde{Y}(t)$  is the solution of  $d\tilde{Y}(t) = c \tilde{Y}(t) dt + \sigma \tilde{Y}(t) dB(t)$  with  $\tilde{Y}(0) = x$ .

However, we have seen in Example 5.11 that the solution  $X(t)$  of  $dX(t) = c X(t) dt + \sigma X(t) dB(t)$  with  $X(0) = x$  is  $X(t) = x \exp[(c - \sigma^2/2)t + \sigma B(t)]$ . For  $\tilde{Y}(t)$  to match  $X(t)$ , the drift in the defining equation for  $Y(t)$  needs to be changed from  $c Y(t)$  to  $(c - \sigma^2/2) Y(t)$ .  $\diamond$

**Theorem 5.10 (Wong-Zakai theorem)** *Let  $X(t)$  and  $X_n(t)$  be a real-valued process defined by the stochastic differential equations*

$$dX(t) = \tilde{a}(X(t), t) dt + b(X(t), t) dB(t), \quad t \in [0, \tau],$$

$$\text{where } \tilde{a}(x, t) = a(x, t) + \frac{1}{2} b(x, t) \frac{\partial b(x, t)}{\partial x} \quad (5.38)$$

and

$$dX_n(t) = a(X_n(t), t) dt + b(X_n(t), t) dB_n(t), \quad t \in [0, \tau], \quad (5.39)$$

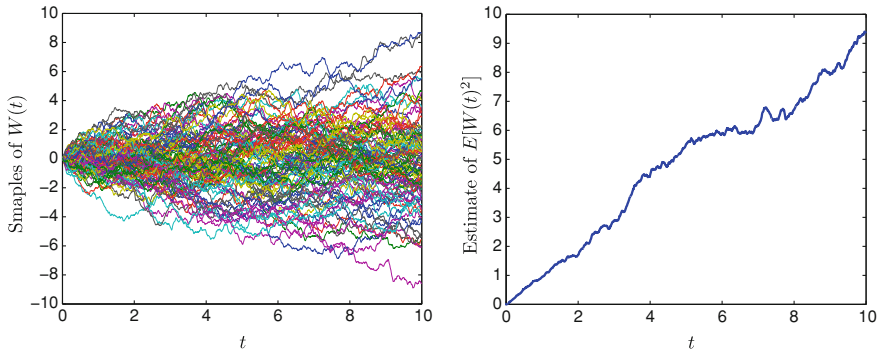
where  $B_n(t)$  is an approximation of  $B(t)$  and initial states  $X(0) = X_n(0)$  are independent of  $B$  and  $B_n$ . If (1) the drift and diffusion coefficients in (5.38) and (5.39) are such that the solutions of these equations exist and are unique in the strong sense, (2)  $B_n(t)$  converges a.s. to  $B(t)$  for all  $t \in [0, \tau]$  as  $n \rightarrow \infty$  and its samples are continuous and of bounded variation on  $[0, \tau]$ , (3)  $B_n$  is uniformly bounded for almost all  $\omega$ , and (4) the samples of  $B_n$  have piecewise continuous derivatives, then  $X_n(t) \rightarrow X(t)$  a.s.,  $n \rightarrow \infty$ , at all  $t \in [0, \tau]$ . If in addition  $B_n(t) \rightarrow B(t)$  uniformly in  $[0, \tau]$ , then  $X_n(t) \rightarrow X(t)$  uniformly in  $[0, \tau]$  a.s. as  $n \rightarrow \infty$  [19].

Approximations  $B_n$  for  $B$  satisfying the requirement of the Wong-Zakai theorem can be constructed simply. For example,  $B_n$  can interpolate linearly between values of  $B$  at the consecutive points of a partition  $p_n = \{t_0, t_1, \dots, t_n\}$ ,  $0 = t_0 < t_1 < \dots < t_n = \tau$ , of  $[0, \tau]$  with mesh  $\Delta(p_n) \rightarrow 0$  as  $n \rightarrow \infty$ . The process  $B_n$  has continuous samples that approach the samples of  $B$  as  $n \rightarrow \infty$  and are of finite variation on compacts, so that  $\int_0^t b(X_n(s), s) dB_n(s)$  can be calculated as a path-by-path Riemann-Stieltjes integral. However,  $B_n$  differs from  $B$  in an essential manner. The Brownian motion is a martingale while  $B_n$  is not. For example,  $B_n(s) = B(t_{i-1}) + \Delta B_i(s - t_{i-1})/\Delta_i$  for  $s \in [t_{i-1}, t_i]$ , where  $\Delta B_i = B(t_i) - B(t_{i-1})$  and  $\Delta_i = t_i - t_{i-1}$ . Since  $B_n(s)$  depends on  $B(t_i)$  is not  $\mathcal{F}_s$ -measurable for  $s < t_i$ , so that it is not a martingale.

Note also that there is no drift correction for equations with additive noise, that is, equations with state independent diffusion coefficients, and that the solution  $X_n(t)$  of the stochastic integral equation

$$X_n(t) = X_n(0) + \int_0^t \left[ a(X_n(s), s) - \frac{1}{2} b(X_n(s), s) \frac{\partial b(X_n(s), s)}{\partial x} \right] ds$$

$$+ \int_0^t b(X_n(s), s) dB_n(s)$$



**Fig. 5.1** Samples of  $W(t)$  (left panel) and an estimate of  $E[W(t)^2]$  (right panel)

driven by colored noise converges in the sense of Wong-Zakai's theorem to the solution  $X(t)$  of the Itô equation (5.27) as  $n \rightarrow \infty$ .

*Example 5.20* Let  $X(t)$ ,  $t \geq 0$ , be a geometrical Brownian motion, so that it satisfies the Itô equation  $dX(t) = cX(t)dt + \sigma X(t)dB(t)$  with initial state  $X(0) = x$ , where  $c, \sigma \in \mathbb{R}$  are constants. The Stratonovich version of this equation is  $\dot{X}(t) = (cX(t) - \sigma^2 X(t)/2) + \sigma X(t)V(t)$  by the Wong-Zakai theorem, where  $V(t)$  is colored stationary Gaussian noise with mean 0 and short memory. For example,  $V(t)$  can be an Ornstein-Uhlenbeck process defined by  $dV(t) = -\alpha V(t)dt + \alpha dB(t)$ ,  $t \geq 0$ , with initial state  $V(0) \sim N(0, \alpha/2)$  that is independent of Brownian motion  $B(t)$ . The stationary covariance function of  $V(t)$  is  $c_v(\tau) = E[V(t+\tau)V(t)] = (\alpha/2) \exp(-\alpha|\tau|)$ .

The solution of the Stratonovich equation can be obtained by classical calculus and is  $X(t) = X(0) \exp[(c - \sigma^2/2)t + \sigma W(t)]$ , where  $W(t) = \int_0^t V(s)ds$  is a Gaussian process with mean 0 and variance  $E[W(t)^2] = t - (1 - \exp(-\alpha t))/\alpha$ . Figure 5.1 shows samples of  $W(t)$  and an estimate of  $E[W(t)^2]$  obtained from 100 independent samples of  $W(t)$  for  $\alpha = 50$ . The samples of  $W(t)$  for this resemble the samples of Brownian motion processes and  $E[W(t)^2] \simeq t$ .  $\diamond$

*Proof* The process  $W(t)$  is Gaussian with mean 0 and covariance function  $E[W(t+\tau)W(t)] = E[W(t)^2] + E\left[\int_t^{t+\tau} V(u)du \int_0^t V(v)dv\right]$  for  $\tau \geq 0$ . The variance of  $W(t)$  is given by  $E[W(t)^2] = \int_{[0,t]^2} c_v(u-v)du dv = 2 \int_0^t (t-\eta) c_v(\eta) d\eta$  by the change of variables  $\xi = v$  and  $\eta = u - v$ , and has the stated expression for  $c_v(\tau) = (\alpha/2) \exp(-\alpha|\tau|)$ . Note that  $\int_{-\infty}^{\infty} c_v(\tau) d\tau = 1$  for every  $\alpha > 0$  so that  $c_v(\tau) \rightarrow \delta(\tau)$  as  $\alpha \rightarrow \infty$ .  $\blacktriangle$

### 5.5.1.2 Semimartingale White Noise

Let  $X(t)$  be the solution of (5.25) and (5.26), where  $Y(t)$  is a semimartingale denoted by  $S(t)$ . As previously, the integrals  $\int_0^t a(X(s-)) ds$  and  $\int_0^t b(X(s-)) dS(s-)$  in (5.26) are Riemann and Itô integrals. We state without proof conditions for the existence and uniqueness of the solution of (5.26).

**Theorem 5.11** *If  $S$  is a semimartingale with  $S(0) = 0$ ,  $X(0)$  is finite and  $\mathcal{F}_0$ -measurable, and the function  $G : [0, \infty) \times \Omega \times \mathbb{R} \rightarrow \mathbb{R}$  is such that (1) the samples  $(t, \omega) \rightarrow G(t, \omega, x)$  are in  $\mathcal{L}$  for a fixed  $x$  and (2)  $|G(t, \omega, x) - G(t, \omega, x')| \leq K(\omega)|x - x'|$  for each  $(t, \omega)$ , where  $K$  is a finite-valued random variable, then the stochastic integral equation*

$$X(t) = X(0) + \int_0^t G(s, \cdot, X(s-)) dS(s) \quad (5.40)$$

*has a unique solution that is a semimartingale ([15], Theorem 6, p. 194).*

An extension of this result is provided by Theorem 5.12 whose statement involves the following concepts. Let  $\mathcal{D}^d$  denote the class of  $\mathbb{R}^d$ -valued stochastic processes whose coordinates are adapted processes with samples in  $\mathcal{D}$ . An operator  $G : \mathcal{D}^d \rightarrow \mathcal{D}$  is said to be functional Lipschitz if, for any processes  $X, Y \in \mathcal{D}^d$ , (1)  $X^{T-} = Y^{T-}$  implies  $G(X)^{T-} = G(Y)^{T-}$  are processes stopped at  $T-$ , where  $T$  denotes a stopping time and (2) there is an increasing finite process  $K(t)$ ,  $t \geq 0$ , such that  $|G(X)(t) - G(Y)(t)| \leq K(t) \|X(t) - Y(t)\|$  a.s. for each  $t \geq 0$  ([15], p. 195).

**Theorem 5.12** *If  $S = (S_1, \dots, S_{d'})$  is a vector of semimartingales,  $S(0) = 0$ ,  $J_i \in \mathcal{D}$ ,  $i = 1, \dots, d$ , and the operators  $G_i^j : \mathcal{D}^d \rightarrow \mathcal{D}$ ,  $i = 1, \dots, d$ ,  $j = 1, \dots, d'$ , are functional Lipschitz, then the system of stochastic integral equations*

$$X_i(t) = J_i(t) + \sum_{j=1}^{d'} \int_0^t G_i^j(X)(s-) dS_j(s), \quad i = 1, \dots, d, \quad (5.41)$$

*has a unique solution in  $\mathcal{D}^d$ . If the processes  $J_i$  are semimartingales, then  $X_i$ ,  $i = 1, \dots, d$ , are also semimartingales ([15], Theorem 7, p. 197).*

**Example 5.21** Let  $dX(t) = b(X(t-), t) dC(t)$  be a stochastic differential equation with the integral form

$$X(t) = X(0) + \int_0^t b(X(s-), s) dC(s) = X(0) + \sum_{k=1}^{N(t)} b(X(T_k-), T_k) Y_k,$$

where  $C(t)$  is a compound Poisson process with jump times  $(T_1, T_2, \dots)$  and iid jumps  $(Y_1, Y_2, \dots)$ . If  $X(0) \in \mathcal{F}_0$ ,  $t \mapsto b(x, t)$  is continuous for each  $x$ , and  $x \mapsto b(x, t)$  satisfies a uniform Lipschitz condition for each  $t \geq 0$ , then the above equation has a unique solution that is a semimartingale.  $\diamond$

*Proof* Theorem 5.11 guarantees the existence and uniqueness of solution  $X$ . The second condition on  $G$  in this theorem becomes the uniform Lipschitz condition  $|b(t, x) - b(t, x')| \leq c |x - x'|$  for each  $t \geq 0$  and a constant  $c > 0$  since  $b$  is a deterministic function. Since  $S = C$ , the samples of  $X$  are constant between consecutive jumps of  $C$  and have jumps  $\Delta X(T_k) = X(T_k) - X(T_k-) = b(X(T_k-), T_k) \Delta C(T_k) = b(X(T_k-), T_k) Y_k$  at the jump times of  $C$  so that  $X(t) = X(T_k-) + b(X(T_k-), T_k) Y_k$ ,  $t \in [T_k, T_{k+1})$ , which gives the stated formula for  $X$ .  $\blacktriangle$

*Example 5.22* Let  $X$  be the solution of  $dX(t) = a(X(t-), t) dt + b(X(t-), t) dS(t)$ , where  $S$  is a semimartingale. If  $X(0) \in \mathcal{F}_0$ , the functions  $t \mapsto a(x, t)$ ,  $b(x, t)$  are continuous for each  $x$ , and the functions  $x \mapsto a(x, t)$ ,  $b(x, t)$  satisfy uniform Lipschitz conditions for each  $t \geq 0$ , then  $X$  is the unique solution of the stochastic integral equation

$$X(t) = X(0) + \int_0^t [a(X(s-), s) + b(X(s-), s) \left[ \frac{ds}{dS(s)} \right]],$$

by Theorem 5.11 since  $(s, S(s))$  is a semimartingale.  $\diamond$

*Example 5.23* Let  $X$  be the process in Example 5.21 and  $g \in C^2(\mathbb{R})$  denote a real-valued function. We have

$$\begin{aligned} g(X(t)) - g(X(0)) &= \int_0^t a(X(s-)) g'(X(s-)) ds \\ &\quad + \int_0^t \int_{\mathbb{R}} [g(X(s-) + y b(X(s-))) - g(X(s-))] \mathcal{M}(ds, dy), \end{aligned} \quad (5.42)$$

where  $\mathcal{M}$  is a Poisson random measure on  $[0, \infty) \times \mathbb{R}$ ,  $E[\mathcal{M}(ds, dy)] = (\lambda ds) dF(y)$ ,  $\lambda > 0$  is the intensity of the Poisson process  $N$ , and  $F$  denotes the distribution of the iid random variables  $Y_k$  ([17], Theorem 4.2.2). Note that (5.42) becomes  $g(C(t)) - g(C(0)) = \int_0^t \int_{\mathbb{R}} [g(C(s-) + y) - g(C(s-))] \mathcal{M}(ds, dy)$  in the special case  $a = 0$  and  $b = 1$ , a result consistent with Example 5.5.  $\diamond$

*Proof* The Itô formula in (5.7) gives

$$\begin{aligned} g(X(t)) - g(X(0)) &= \int_{0+}^t g'(X(s-)) dX(s) + \frac{1}{2} \int_{0+}^t g''(X(s-)) d[X, X]^c(s) \\ &\quad + \sum_{0 < s \leq t} [g(X(s) - g(X(s-)) - g'(X(s-)) \Delta X(s)] \end{aligned}$$

so that

$$g(X(t)) - g(X(0)) = \int_{0+}^t g'(X(s-)) a(X(s-)) ds + \sum_{0 < s \leq t} [g(X(s)) - g(X(s-))]. \quad (5.43)$$

since  $\Delta X(s) = b(X(s-)) \Delta C(s)$ ,  $[X, X]^c = 0$ , and  $\int_{0+}^t g'(X(s-)) b(X(s-)) dC(s) = \sum_{0 < s \leq t} g'(X(s-)) b(X(s-)) \Delta C(s) = \sum_{0 < s \leq t} g'(X(s-)) \Delta X(s)$ . This formula, the integral form  $\int_0^t \int_{\mathbb{R}} [g(X(s-) + y b(X(s-)) - g(X(s-))] \mathcal{M}(ds, dy)$  of  $\sum_{0 < s \leq t} [g(X(s)) - g(X(s-))]$ , and  $g(X(s)) = g(X(s-)) + b(X(s-)) \Delta C(s)$  give (5.42).  $\blacktriangle$

*Example 5.24* Let  $X(t)$  be the solution of  $dX(t) = -\alpha X(t-) dt + dC(t)$ ,  $t \geq 0$ , starting at  $X(0) = X_0$ , where  $\alpha > 0$  is a constant,  $C(t) = \sum_{k=1}^{N(t)} Y_k$  is a compound Poisson process,  $N(t)$  denotes a Poisson process with intensity  $\lambda > 0$ , and  $\{Y_k\}$  are iid random variables. It is assumed that  $X_0$  is independent of  $C(t)$  and that  $X_0$  and  $Y_1$  have finite moments of any order. The moments  $\mu(q; t) = E[X(t)^q]$  of  $X(t)$  satisfy the ordinary differential equation

$$\dot{\mu}(q; t) = -\alpha q \mu(q; t) + \lambda \sum_{k=1}^q \frac{q!}{k! (q-k)!} \mu(q-k; t) E[Y_1^k], \quad t \geq 0, \quad (5.44)$$

with initial condition  $\mu(q; 0) = E[X_0^q]$  and the convention  $\mu(r; t) = 0$  for  $r < 0$ , where  $q \geq 1$  is an integer.  $\diamond$

*Proof* The version of Itô's formula in (5.43) with  $g(x) = x^q$  gives

$$\begin{aligned} X(t)^q - X(0)^q &= -\alpha q \int_{0+}^t X(s-)^q ds + \sum_{0 < s \leq t} [(X(s-) + \Delta X(s))^q - X(s-)^q] \\ &= -\alpha q \int_{0+}^t X(s-)^q ds + \sum_{0 < s \leq t} \left[ \sum_{k=1}^q \frac{q!}{k! (q-k)!} X(s-)^{q-k} \Delta X(s)^k \right]. \end{aligned} \quad (5.45)$$

The expectation of the left side of this equation is  $\mu(q; t) - \mu(q; 0)$ . The expectation of the first term on the right side is  $-\alpha q \int_{0+}^t \mu(q; s) ds$  by Fubini's theorem since  $X$  is integrable and measurable in both arguments. We have  $E[X(s-)^q] = E[X(s)^q] = \mu(q; s)$  since  $X(s-) = X(s)$  with probability 1 at an arbitrary but fixed  $s$ . The latter statement holds since  $X(s-) = \lim_{u \uparrow s} X(u)$  differs from  $X(s)$  if and only if  $C$  has a jump at  $s$ . However, the probability that  $C$  has at least a jump in  $(u, s]$ ,  $u < s$ , is  $P(N(s-u) \geq 1) = 1 - \exp(-\lambda(s-u))$  and approaches 0 as  $u \uparrow s$  implying that  $X(s-)$  and  $X(s)$  can only differ on a set of measure zero. The derivatives with respect to  $t$  of the terms on the left side and the first term on the right side of (5.45) are  $\dot{\mu}(q; t)$  and  $-\alpha q \mu(q; t)$ .

Consider now the second term on the right side of (5.45). The expectation of the increment of this term in the time interval  $(t, t + \Delta t]$  is

$$\begin{aligned}
& E \left[ \sum_{t < s \leq t + \Delta t} \left[ \sum_{k=1}^q \frac{q!}{k! (q-k)!} X(s-)^{q-k} \Delta X(s)^k \right] \right] \\
&= E \left[ \sum_{n=0}^{\infty} \left( \sum_{k=1}^q \frac{q!}{k! (q-k)!} X(T_n-)^{q-k} \Delta X(T_n)^k \right) \mid N(\Delta t) = n \right] P(N(\Delta t) = n),
\end{aligned}$$

where  $\{T_n\}$  denote jump times of  $C$  in the time interval  $(t, t + \Delta t]$  and  $P(N(\Delta t) = n) = (\lambda \Delta t)^n \exp(-\lambda \Delta t) / n!$ ,  $n \geq 0$ . The above expectations can be calculated term by term since  $C$  has a finite number of jumps in any finite time interval with probability 1. Under the assumption that  $X$  has finite moments of any order, the term corresponding to  $n = 0$  is zero since  $\Delta X(s) = 0$  for all  $s \in (t, t + \Delta t]$  and the terms corresponding to  $n \geq 1$  are of order  $O(\Delta t)^n$ , so that the term

$$\begin{aligned}
& E \left[ \sum_{t < s \leq t + \Delta t} \left[ \sum_{k=1}^q \frac{q!}{k! (q-k)!} X(s-)^{q-k} \Delta X(s)^k \right] \right] \\
&= \sum_{k=1}^q \frac{q!}{k! (q-k)!} E[X(T_1-)^{q-k} \Delta X(T_1)^k] \lambda \Delta t \exp(-\lambda \Delta t) + O(\Delta t)^2 \\
&= \sum_{k=1}^q \frac{q!}{k! (q-k)!} \mu(q-k; t) E[Y_1^k] \lambda \Delta t \exp(-\lambda \Delta t) + O(\Delta t)^2,
\end{aligned}$$

scaled by  $\Delta t$  converges to the second term on the right side of (5.44) as  $\Delta t \rightarrow 0$ .  $\blacktriangle$

### 5.5.2 Tanaka's Formula

Itô's formula cannot be applied to  $B \mapsto |B|$  since this mapping is not in  $C^2(\mathbb{R})$ . Mappings of these type are needed to solve locally a class of deterministic partial differential equations with mixed boundary conditions (Sect. 5.5.3). The extension of Itô's formula to the mapping  $B \mapsto |B|$  is referred to as Tanaka's formula.

**Definition 5.5** The local time process is

$$L(t) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t 1(-\varepsilon < B(s) < \varepsilon) ds, \quad (5.46)$$

where the limit exists in  $L^2$  and a.s. ([3], Chap. 7).

**Theorem 5.13 (Tanaka's formula)** *If  $B$  is a Brownian motion, then*

$$\begin{aligned}
|B(t)| - |B(0)| &= \int_0^t \text{sign}(B(s)) dB(s) + \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t 1(-\varepsilon < B(s) < \varepsilon) ds \\
&= \hat{B}(t) + L(t) \quad \text{a.s.},
\end{aligned} \quad (5.47)$$



where the limit is in  $L^2$ ,  $\hat{B}$  is a Brownian motion, and  $\text{sign}(x)$  is  $-1$ ,  $0$ , and  $1$  for  $x < 0$ ,  $x = 0$ , and  $x > 0$ , respectively.

*Proof* For a proof of Tanaka's formula see [3] (Sect. 7.3) and [12] (Sect. 8.6). We only note that  $|B|$  is a semimartingale since (1)  $L$  is a continuous increasing process that is adapted to the natural filtration of  $B$  and (2)  $\hat{B}$  is a square integrable martingale with continuous samples and quadratic variation  $[\hat{B}, \hat{B}](t) = \int_0^t (\text{sign}(B(s)))^2 d[B, B](s) = t$  a.s. Hence,  $\hat{B}$  is a Brownian motion (Example 5.8). Note that formal application of Itô's formula to the mapping  $B \mapsto g(B) = |B|$  gives

$$|B(t)| - |B(0)| = \int_0^t \text{sign}(B(s)) dB(s) + \int_0^t \delta(B(s)) ds$$

since  $g'(x) = \text{sign}(x)$  and  $g''(x) = 2\delta(x)$ . The first integral in this equation is  $\hat{B}(t)$  in Tanaka's formula while the second integral corresponds to  $L(t)$ .  $\blacktriangle$

Following are two extensions of Tanaka's formula. The first and second extensions involve Brownian motion reflected at zero and at two thresholds, respectively.

**Theorem 5.14** *If  $g \in C^2(\mathbb{R})$  and  $\tilde{L}(t) = \int_0^t g'(|B(s)|) dL(s) = g'(0)L(t)$ , then*

$$\begin{aligned} g(|B(t)|) - g(|B(0)|) \\ = \int_0^t g'(|B(s)|) \text{sign}(B(s)) dB(s) + \frac{1}{2} \int_0^t g''(|B(s)|) ds + \tilde{L}(t). \end{aligned} \quad (5.48)$$

*Proof* Since  $g \in C^2(\mathbb{R})$  and the  $\mathbb{R}^2$ -valued process  $X = (\hat{B}, L)$  is a semimartingale, the Itô formula in (5.15) applied to mapping  $X \mapsto g(X) = g(\hat{B} + L)$  gives

$$\begin{aligned} g(|B(t)|) - g(|B(0)|) &= \int_0^t g'(|B(s)|) d(\hat{B} + L)(s) + \frac{1}{2} \int_0^t g''(|B(s)|) ds \\ &= \int_0^t g'(|B(s)|) d\hat{B}(s) + \frac{1}{2} \int_0^t g''(|B(s)|) ds + \int_0^t g'(|B(s)|) dL(s) \\ &= \int_0^t g'(|B(s)|) d\hat{B}(s) + \frac{1}{2} \int_0^t g''(|B(s)|) ds + g'(0)L(t), \end{aligned}$$

which is the formula in (5.48). An alternative proof is in [7] (Sect. 6.2.3.1).  $\blacktriangle$

**Example 5.25** Let  $u$  be the solution of Schrödinger equation  $(1/2)u''(x) + qu(x) = 0$ ,  $x \in D = (0, l)$ , with the Dirichlet and Neumann boundary conditions  $u'(0) = \alpha$  and  $u(l) = \beta$ , where  $\alpha, \beta$  are some constants. The local solution of this equation is

$$u(x) = \beta E^x[e^{qT}] - \alpha E^x\left[\int_0^T e^{qs} dL(s)\right],$$

where  $T = \inf\{t > 0 : |B(t)| \notin D\}$ ,  $B(0) = x \in D$ , and  $E^x[\cdot] = E[\cdot | B(0) = x]$ .

The exact solution of the Schrödinger equation is

$$u(x) = \frac{\beta - (\alpha/\sqrt{2q}) \sin(\sqrt{2q}l)}{\cos(\sqrt{2q}l)} \cos(\sqrt{2q}x) + \frac{\alpha}{\sqrt{2q}} \sin(\sqrt{2q}x)$$

for  $q > 0$ . The local solution of this equation is based on an  $\mathbb{R}^2$ -valued process  $X$  with coordinates  $X_1 = |B|$  and  $X_2$  defined by  $dX_2(t) = q X_2(t) dt$  with  $X_2(0) = 1$ . Estimates of  $u$  for  $\alpha = 1$ ,  $\beta = 2$ ,  $q = 1$ , and  $l = 1$  based on  $n = 500$  independent samples of  $X$  generated at a time step of  $\Delta t = 0.0005$  are in error by less than 2%. The error has been calculated with respect to the exact solution.  $\diamond$

*Proof* Itô's formula applied to mapping  $X \mapsto u(|B(t)|) X_2(t) = u(\hat{B}(t) + L(t)) X_2(t)$  gives

$$\begin{aligned} u(|B(t)|) X_2(t) - u(x) &= \int_0^t u'(\hat{B}(s) + L(s)) X_2(s) d(\hat{B}(s) + L(s)) \\ &+ \int_0^t u(\hat{B}(s) + L(s)) dX_2(s) + \frac{1}{2} \int_0^t u''(\hat{B}(s) + L(s)) X_2(s) ds, \end{aligned}$$

so that

$$\begin{aligned} E^x [u(|B(T)|) X_2(T)] - u(x) &= E^x \left[ \int_0^T u'(|B(s)|) X_2(s) dL(s) \right] \\ &+ q E^x \left[ \int_0^T u(|B(s)|) X_2(s) ds \right] + \frac{1}{2} E^x \left[ \int_0^T u''(|B(s)|) X_2(s) ds \right] \end{aligned}$$

by averaging and using  $T$  in place of  $t$  ([14], Chap. 9). Since  $|B(s)|$  is in  $D$  for  $s < T$ , then  $(1/2)u''(|B(s)|) + q u(|B(s)|) = 0$  so that

$$E^x [u(|B(T)|) X_2(T)] - u(x) = u'(0) E^x \left[ \int_0^T X_2(s) dL(s) \right],$$

which gives  $u(x)$  since  $u(|B(T)|) = \beta$ ,  $X_2(s) = e^{qs}$ , and  $u'(0) = \alpha$ .  $\blacktriangle$

Let  $r$  be a periodic function with period  $2(b-a)$ ,  $a < b$ , defined by  $r(x) = |x-a|$  for  $|x-a| \leq b-a$ . The process  $r(B)$  has the range  $[0, b-a]$ , and is referred to as Brownian motion reflected at two thresholds. Let  $x_k(a) = a + 2k(b-a)$  and  $x_k(b) = b + 2k(b-a)$ ,  $k \in \mathbb{Z}$ , be the collection of points on the real line where  $r(x)$  is equal to 0 and  $b-a$ , respectively. For  $\varepsilon \in (0, (b-a)/2)$  define the intervals  $I_{k,\varepsilon}(a) = \{x : |x - x_k(a)| \leq \varepsilon\}$  and  $I_{k,\varepsilon}(b) = \{x : |x - x_k(b)| \leq \varepsilon\}$  centered on  $x_k(a)$  and  $x_k(b)$ , respectively, and set  $I(a) = \cup_{k=-\infty}^{\infty} (x_k(a), x_k(b))$  and  $I(b) = \cup_{k=-\infty}^{\infty} (x_k(b), x_{k+1}(a))$ .

**Theorem 5.15** *If  $g \in C^2(\mathbb{R})$ , then*

$$\begin{aligned} g(r(B(t))) - g(r(B(0))) &= \int_0^t g'(r(B(s))) (1(B(s) \in I(a)) - 1(B(s) \in I(b))) dB(s) \\ &+ \sum_{k=-\infty}^{\infty} [g'(0) L(t; x_k(a)) - g'(b-a) L(t; x_k(b))] + \frac{1}{2} \int_0^t g''(r(B(s))) ds, \end{aligned} \quad (5.49)$$

where  $L(t; x_k(a)) = \lim_{\varepsilon \downarrow 0} (1/2\varepsilon) \int_0^t 1(B(s) \in I_{k,\varepsilon}(a)) ds$  and  $L(t; x_k(b))$  is  $L(t; x_k(a))$  with  $I_{k,\varepsilon}(b)$  in place of  $I_{k,\varepsilon}(a)$ .

*Proof* For proof see [7] (Sect. 6.2.3.2). Tanaka's formula and its previous extension cannot be used to prove (5.49) since they deal with Brownian motion processes reflected at zero. ▲

### 5.5.3 Random Walk Method

We show that solutions of some deterministic partial differential equations can be obtained at arbitrary points in their domains of definition directly, rather than extracting them from field solutions. The method delivering these local solutions uses samples of diffusion processes, and is referred to as the random walk method.

The class of partial differential equations admitting local solutions has the form

$$\begin{aligned} \frac{\partial u(x, t)}{\partial t} &= \sum_{i=1}^d \alpha_i(x) \frac{\partial u(x, t)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d \beta_{ij}(x) \frac{\partial^2 u(x, t)}{\partial x_i \partial x_j} \\ &+ q(x, t) u(x, t) + p(x, t), \quad (x, t) \in D \times (0, \infty), \end{aligned} \quad (5.50)$$

where  $D$  is an open subset of  $\mathbb{R}^d$ ,  $\alpha_i, \beta_{ij}$  are real-valued functions defined on  $D \subset \mathbb{R}^d$ ,  $d \geq 1$  is an integer, and  $q, p$  denote real-valued functions defined on  $D \times [0, \infty)$ . The solution  $u : D \times (0, \infty) \rightarrow \mathbb{R}$  of (5.50) depends on boundary and initial conditions. If  $q$  and  $p$  depend on only  $x$  and  $\partial u / \partial t = 0$ , (5.50) is a partial differential equations involving only spatial coordinates. Note that (5.50) can be a Poisson equation, heat or any other transport equation, or a Schrödinger equation depending on the coefficients  $\alpha_i(x)$  and  $\beta_{ij}(x)$  and the functions  $q(x, t)$  and  $p(x, t)$ .

#### 5.5.3.1 Dirichlet Boundary Value Problem ( $q = 0$ )

Let  $u(x, t)$  be the solution of (5.50) with  $q = 0$ , satisfying the initial and boundary conditions  $u(x, 0) = \eta(x)$ ,  $x \in D$ , and  $u(x, t) = \xi(x, t)$ ,  $x \in \partial D$ ,  $t > 0$ .

It is assumed that (1) the matrix  $\beta(x) = \{\beta_{ij}(x)\}$  in (5.50) is symmetric and positive definite admitting the representation  $\beta(x) = b(x) b(x)'$ , (2) the stochastic differential equation

$$\begin{cases} dX(s) &= a(X(s)) ds + b(X(s)) dB(s) \\ dX_{d+1}(s) &= -ds \end{cases} \quad (5.51)$$

has a unique strong solution  $\tilde{X}(s) = (X(s), X_{d+1}(s)) \in \mathbb{R}^{d+1}$ ,  $s \geq 0$ , where  $B$  denotes an  $\mathbb{R}^d$ -valued Brownian motion,  $a_i(x) = \alpha_i(x)$ ,  $i = 1, \dots, d$ , are given by (5.50), and  $b(x)$  is in the representation of  $\beta(x)$ , and (3) the boundaries of  $D$  are regular, the partial derivatives of  $u$  in (5.50) are bounded in  $D_t = D \times (0, t)$ ,  $t > 0$ , the functions  $\eta$  and  $\xi$  are continuous in their arguments, and  $p$  is Hölder continuous in  $D$ , that is, there exist constants  $c, \alpha \in (0, \infty)$  such that  $|p(x, t) - p(x', t)| \leq c \|x - x'\|^\alpha$ , where  $x, x' \in D$  ([4], p. 133).

**Definition 5.6** The generator of the diffusion process  $\tilde{X} = (X, X_{d+1})$  in (5.51) is the limit

$$\mathcal{A}[g(x, t)] = \lim_{s \downarrow 0} \frac{E^{(x, t)}[g(\tilde{X}(s))] - g(x, t)}{s}, \quad (5.52)$$

where  $g$  is a real-valued function defined on  $\mathbb{R}^{d+1}$  with continuous second order partial derivatives in  $x \in \mathbb{R}^d$  and continuous first order partial derivative in  $s \geq 0$  and  $E^{(x, t)}$  denotes expectation conditional on  $\tilde{X}(0) = (x, t)$ ,  $x \in D$  and  $t > 0$ .

**Theorem 5.16** Under the stated assumptions, the local solution of (5.50) with  $q = 0$  is

$$\begin{aligned} u(x, t) &= E^{(x, t)}[u(\tilde{X}(\tilde{T}))] + E^{(x, t)}\left[\int_0^{\tilde{T}} p(\tilde{X}(s)) ds\right] \\ &= E^{(x, t)}[\eta(X(t)) \mid \tilde{T} = t] P(\tilde{T} = t) + E^{(x, t)}[\xi(\tilde{X}(t - \tilde{T})) \mid \tilde{T} < t] P(\tilde{T} < t) \\ &\quad + E^{(x, t)}\left[\int_0^{\tilde{T}} p(\tilde{X}(s)) ds\right], \end{aligned} \quad (5.53)$$

where  $\tilde{X} = (X, X_{d+1})$  and  $\tilde{T} = \inf\{s > 0 : \tilde{X}(s) \notin D_t, x \in D, t > 0\}$ .

*Proof* We present the main idea of the proof. Technical arguments needed to support some of the statements in our discussion can be found in [3] (Chap. 6), [4] (Chap. 4), [7] (Sect. 6.2), and [14] (Chap. 9).

The Itô formula in (5.15) applied to the mapping  $\tilde{X} \mapsto g(\tilde{X})$  gives

$$\begin{aligned} E^{(x, t)}[g(\tilde{X}(s))] - g(\tilde{X}(0)) &= \sum_{i=1}^{d+1} E^{(x, t)}\left[\int_0^s \frac{\partial g(\tilde{X}(\sigma))}{\partial \tilde{x}_i} d\tilde{X}_i(\sigma)\right] \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d E^{(x, t)}\left[\int_0^s \left(b(X(\sigma)) b(X(\sigma))'\right)_{ij} \frac{\partial^2 g(\tilde{X}(\sigma))}{\partial \tilde{x}_i \partial \tilde{x}_j} d\sigma\right] \end{aligned}$$

by averaging, where  $g$  is as in (5.52),  $\tilde{x}_i$  is the coordinate  $i = 1, \dots, d, d+1$  of  $\tilde{x} = (x, x_{d+1})$ , and  $\tilde{x}_i = x_i$  for  $i = 1, \dots, d$ . An alternative form of this equation is

$$\begin{aligned}
E^{(x,t)} \left[ g(\tilde{X}(s)) \right] - g(x, t) &= E^{(x,t)} \left[ \int_0^s \left( \sum_{i=1}^d \frac{\partial g(\tilde{X}(\sigma))}{\partial x_i} a_i(X(\sigma)) - \frac{\partial g(\tilde{X}(\sigma))}{\partial x_{d+1}} \right) d\sigma \right. \\
&\quad \left. + \frac{1}{2} \sum_{i,j=1}^d \int_0^s \left( b(X(\sigma)) b(X(\sigma))' \right)_{ij} \frac{\partial^2 g(\tilde{X}(\sigma))}{\partial x_i \partial x_j} d\sigma \right] \quad (5.54)
\end{aligned}$$

since the processes  $\int_0^s \left[ \partial g(\tilde{X}(\sigma)) / \partial \tilde{x}_i \right] b_{ij}(\tilde{X}(\sigma)) dB_j(\sigma)$ ,  $i, j = 1, \dots, d$ , are martingales starting at zero. For a small  $s > 0$ , the argument of the expectation on the right side of the above equation divided by  $s$  can be approximated by

$$\begin{aligned}
&\sum_{i=1}^d \frac{\partial g(\tilde{X}(\theta(\omega)s, \omega))}{\partial x_i} a_i(X(\theta(\omega)s, \omega)) - \frac{\partial g(\tilde{X}(\theta(\omega)s, \omega))}{\partial x_{d+1}} \\
&+ \frac{1}{2} \sum_{i,j=1}^d \left( b(X(\theta(\omega)s, \omega)) b(X(\theta(\omega)s, \omega))^T \right)_{ij} \frac{\partial^2 g(\tilde{X}(\theta(\omega)s, \omega))}{\partial x_i \partial x_j}
\end{aligned}$$

for each  $\omega \in \Omega$ , where  $\theta(\omega) \in [0, 1]$ . Since the drift and diffusion coefficients of  $\tilde{X}$  satisfy the conditions in Theorem 5.7,  $\tilde{X}$  has continuous sample paths, and  $g$  has continuous partial derivatives of order 1 and order 2 with respect to  $x_{d+1}$  and  $x_i$ ,  $i = 1, \dots, d$ , respectively, the limit of the above function as  $s \downarrow 0$  gives

$$\mathcal{A} = -\frac{\partial}{\partial x_{d+1}} + \sum_{i=1}^d a_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d (b b^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j}, \quad (5.55)$$

so that

$$E^{(x,t)} \left[ g(\tilde{X}(s)) \right] - g(x, t) = E^{(x,t)} \left[ \int_0^s \mathcal{A} \left[ g(\tilde{X}(\sigma)) \right] d\sigma \right]. \quad (5.56)$$

Recall that calculations similar to those used here to establish the generator  $\mathcal{A}$  of  $\tilde{X}$  were performed in Example 5.7 to show that the generator of a Brownian motion is proportional to the Laplace operator.

Note that (5.56) holds with  $u$  in place of  $g$  and that the local solution in (5.53) results from (5.56) for  $s = \tilde{T}$  since  $\tilde{X}(s) \in D_t$  for  $s < \tilde{T}$  and  $\tilde{X}(\tilde{T})$  is on the boundary of  $D_t$  so that  $\mathcal{A} \left[ u(\tilde{X}(s)) \right] = -p(\tilde{X}(s))$  and  $u(\tilde{X}(\tilde{T})) = \xi(\tilde{X}(t - \tilde{T}))$  for  $\tilde{T} < t$  and  $\eta(X(t))$  for  $\tilde{T} = t$ . That (5.56) can be used for  $s = \tilde{T}$  follows from the Dynkin formula ([14], Theorem 7.4.1).  $\blacktriangle$

**Theorem 5.17** *Under the conditions in Theorem 5.16, the local solution of the time invariant version of (5.50) with  $q = 0$  and  $u(x) = \xi(x)$ ,  $x \in \partial D$ , is*

$$u(x) = E^x \left[ \xi(X(T)) \right] + E^x \left[ \int_0^T p(X(\sigma)) d\sigma \right], \quad (5.57)$$

where  $X$  is an  $\mathbb{R}^d$ -valued diffusion process satisfying the first  $d$  equations in (5.51) and  $T = \inf\{s > 0 : X(s) \notin D\}$ .

*Proof* Arguments similar to those used to derive (5.53) can be used to find (5.57). The generator of  $X$  is

$$\mathcal{A} = \sum_{i=1}^d a_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d (b(x) b(x)')_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \quad (5.58)$$

so that

$$E^x [g(X(s))] - g(x) = E^x \left[ \int_0^s \mathcal{A} [g(X(\sigma))] d\sigma \right] \quad (5.59)$$

by Itô's formula applied to  $g(X)$  for  $s \in [0, T)$ .  $\blacktriangle$

Estimates of solutions  $u(x, t)$  and  $u(x)$  can be calculated from (5.53) and (5.57) using samples of  $\tilde{X}$  and  $X$ , respectively. The estimates are satisfactory if the sample size is sufficiently large and the numerical scheme used to generate samples of  $\tilde{X}$  and  $X$  is accurate. The accuracy of this scheme depends strongly on the time step  $\Delta t$  that needs to be related to the gradient of the coefficients in (5.50), as illustrated by numerical experiments in [9].

*Example 5.26* The local solution for

$$\sum_{i=1}^2 \alpha_i(x) \frac{\partial u(x)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^2 \beta_{ij}(x) \frac{\partial^2 u(x)}{\partial x_i \partial x_j} = 0, \quad x \in D \subset \mathbb{R}^2,$$

with the boundary condition  $u(x) = \xi(x)$ ,  $x \in \partial D$ , is  $u(x) = E^x [u(X(T))]$  by (5.57), where  $X$  is an  $\mathbb{R}^2$ -valued diffusion process defined by the first  $d = 2$  equations in (5.51),  $T$  is as in Theorem 5.17 such that  $E^x [T] < \infty$ , and  $X(0) = x \in D$ .

Estimates of  $u(x)$  have been calculated for  $D = (-2, 2) \times (-1, 1)$ ,  $\xi(x) = 1$ ,  $x \in (-2, 2) \times \{-1\}$ , and  $\xi(x) = 0$  on the other boundaries of  $D$ ,  $\alpha_1(x) = 1$ ,  $\alpha_2(x) = 2$ ,  $\beta_{11}(x) = 1.25$ ,  $\beta_{12}(x) = \beta_{21}(x) = 1.5$ , and  $\beta_{22}(x) = 4.25$ . The estimates are based on 1000 independent samples of  $X$  defining by

$$\begin{cases} dX_1(t) = dt + dB_1(t) + (1/2) dB_2(t), \\ dX_2(t) = 2 dt + (1/2) dB_1(t) + 2 dB_2(t), \end{cases}$$

and generated with a time step of  $\Delta t = 0.001$ , where  $B_1$  and  $B_2$  are independent Brownian motions. They are 0.2720, 0.2650, and 0.2222 at  $x = (0.5, 0)$ ,  $(1.0, 0)$ , and  $(1.5, 0)$ , and differ from the finite element solution by less than 5%.  $\diamond$

*Example 5.27* Consider a bar in torsion with a multiply connected cross section  $D$ , external boundary  $\partial D_1$ , and interior boundaries  $\partial D_r$ ,  $r = 2, \dots, m$ , delineating cavities. The Prandtl stress function  $u(x)$  for this bar is the solution of  $\Delta u(x) =$

$-2G\theta$ ,  $x \in D \subset \mathbb{R}^2$ , where  $G$  is the modulus of elasticity in shear and  $\theta$  denotes the angle of twist.

The local solution is

$$u(x) = E^x[u(B(T))] - \frac{1}{2} E^x \left[ \int_0^T \Delta u(B(s)) ds \right] = \sum_{r=2}^m k_r P(B(T) \in \partial D_r) + G\beta E^x[T]$$

for the boundary conditions  $u(x) = k_r$ ,  $x \in \partial D_r$ ,  $r = 1, \dots, m$ , where  $k_r$  are constants. One of the constants  $k_r$  is arbitrary and can be set to zero, for example  $k_1 = 0$  ([1], Sect. 7-6). If the Brownian motion  $B$  starts at a point on the boundary of  $D$ , for example,  $x \in \partial D_r$ , and  $D$  is regular, then  $P(B(T) \in \partial D_r) = 1$ ,  $P(B(T) \in \partial D_q) = 0$ ,  $q \neq r$ , and  $E^x[T] = 0$  so that  $u(x) = k_r$ , that is, the local solution satisfies the boundary conditions exactly [5, 6].

If  $D$  is a simply connected set, then  $p_r(x) = 0$  for  $r = 2, \dots, m$  so that  $u(x) = G\theta E^x[T]$ . This solution has been applied to an elliptical cross section  $D = \{x : (x_1/a_1)^2 + (x_2/a_2)^2 < 1\}$  with  $a_1 = 5$ ,  $a_2 = 3$ , and  $G\theta = 1/2$ . For 100 samples of  $B$  generated with a time step  $\Delta t = 0.001$  the largest error of the estimated Prandtl function is nearly 33%. The error decreases to 2% if the sample size is increased to 1000. Errors have been calculated relative to the exact solution  $u(x) = -a_1^2 a_2^2 G\beta (x_1^2/a_1^2 + x_2^2/a_2^2 - 1)/(2(a_1^2 + a_2^2))$  ([1], Sect. 7-4).  $\diamond$

*Example 5.28* Let  $u$  satisfy the heat equation

$$\frac{\partial u(x, t)}{\partial t} = \alpha^2 \sum_{i=1}^d \frac{\partial^2 u(x, t)}{\partial x_i^2} \quad (5.60)$$

with initial and boundary conditions giving the functions  $u(x, 0)$  for  $x \in D$  and  $u(x, t)$  for  $x \in \partial D$  and  $t > 0$ , respectively, where  $D$  is an open bounded set in  $\mathbb{R}^d$  and  $\alpha > 0$  is a constant. The local solution of (5.60) is

$$u(x, t) = E^{(x, t)}[u(\tilde{X}(\tilde{T}))], \quad (5.61)$$

where  $d\tilde{X}_i(s) = \sqrt{2}\alpha dB_i(s)$ ,  $i = 1, \dots, d$  and  $d\tilde{X}_{d+1}(s) = -ds$  by (5.51).

*Proof* Note that (5.60) is a special case of (5.50) with  $\alpha_i = 0$ ,  $\beta_{ij} = 2\alpha^2 \delta_{ij}$ ,  $q = 0$ , and  $p = 0$ . The generator of  $\tilde{X}$  is  $\mathcal{A} = -\partial/\partial x_{d+1} + \alpha^2 \Delta$  and the steady-state solution  $u_s(x) = \lim_{t \rightarrow \infty} u(x, t)$  satisfies the Laplace equation  $\Delta u_s = 0$ .  $\blacktriangle$

### 5.5.3.2 Dirichlet Boundary Value Problem ( $q \neq 0$ )

The local solution for this general form of (5.50) involves the Feynman-Kac functional. Results in the previous section can be viewed as a special case of those discussed here.

**Theorem 5.18** Let  $\rho, \zeta : \mathbb{R}^d \rightarrow \mathbb{R}$  be functions such that  $\rho$  has a compact support and continuous second order partial derivatives and  $\zeta$  is a bounded Borel measurable

function. Let  $X$  be an  $\mathbb{R}^d$ -valued diffusion process defined by  $dX(t) = a(X(t)) dt + b(X(t)) dB(t)$ ,  $t \geq 0$ , that has a unique solution that belongs a.s. to the support of  $\rho$ . The Feynman-Kac functional,

$$v(x, t) = E^x \left[ \rho(X(t)) e^{\int_0^t \zeta(X(u)) du} \right], \quad (5.62)$$

is the solution of the partial differential equation

$$\frac{\partial v}{\partial t} = \mathcal{A} [v] + \zeta v \quad (5.63)$$

with  $v(x, 0) = \rho(x)$  and some boundary conditions, where  $\mathcal{A}$  denotes the generator of  $X(t)$ .

*Proof* A sketch of the proof can be found in [7] (Sect. 6.2.2). Additional technical considerations are in [14] (Theorem 8.2.1).  $\blacktriangle$

**Theorem 5.19** The local solution of (5.50) is

$$\begin{aligned} u(x, t) = & E^{(x,t)} \left[ u(\tilde{X}(\tilde{T})) \exp \left( \int_0^{\tilde{T}} q(\tilde{X}(\sigma)) d\sigma \right) \right] \\ & + E^{(x,t)} \left[ \int_0^{\tilde{T}} p(\tilde{X}(s)) \exp \left( \int_0^s q(\tilde{X}(\sigma)) d\sigma \right) ds \right], \end{aligned} \quad (5.64)$$

for  $x \in D$  and  $t > 0$ , where  $\tilde{X}$  is given by (5.51),  $\tilde{T}$  is as in (5.53), and  $\rho$ ,  $\zeta$ , and  $q$  are such that (5.62) and (5.63) apply.

*Proof* Consider an  $\mathbb{R}^{d+2}$ -valued diffusion process  $(\tilde{X}, X_{d+2}) = (X, X_{d+1}, X_{d+2} = Z)$  defined by the stochastic differential equation

$$\begin{cases} dX(s) &= a(X(s)) ds + b(X(s)) dB(s), \\ dX_{d+1}(s) &= -ds, \\ dZ(s) &= q(X(s), s) Z(s) ds, \end{cases}$$

for  $s \geq 0$  with initial conditions  $X(0) = x \in D$ ,  $X_{d+1}(0) = t > 0$ , and  $X_{d+2}(0) = Z(0) = 1$ . The generator of  $(\tilde{X}, Z) = (X, X_{d+1}, X_{d+2} = Z)$  is

$$\mathcal{A}^* = -\frac{\partial}{\partial x_{d+1}} + \sum_{i=1}^d a_i \frac{\partial}{\partial x_i} + qz \frac{\partial}{\partial x_{d+2}} + \frac{1}{2} \sum_{i,j=1}^d (b b^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j},$$

so that

$$\begin{aligned} \mathcal{A}^* [g(x, s, z)] = & z \left[ -\frac{\partial}{\partial x_{d+1}} + \sum_{i=1}^d a_i(x) \frac{\partial}{\partial x_i} + q(x, s) \right. \\ & \left. + \frac{1}{2} \sum_{i,j=1}^d (b(x) b(x)^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \right] \psi(x, x_{d+1}). \end{aligned}$$



for  $g(x, x_{d+1}, x_{d+2} = z) = \psi(x, x_{d+1})z$  assumed to have continuous first order partial derivatives in  $s$  and  $z$  and continuous second order partial derivatives in  $x$ .

For  $g(\tilde{X}, Z) = u(\tilde{X})Z$ , the Itô formula gives

$$E^{(x,t)} \left[ u(\tilde{X}(\tilde{T})) Z(\tilde{T}) \right] - u(\tilde{X}(0)) Z(0) = E^{(x,t)} \left[ \int_0^{\tilde{T}} \mathcal{A}^* \left[ u(\tilde{X}(s)) Z(s) \right] ds \right]$$

by averaging. This formula yields (5.64) since  $(\tilde{X}, Z)$  starts at  $(x, t, 1)$  so that  $u(\tilde{X}(0))Z(0) = u(x, t)$ ,  $\tilde{X}(s) \in D_t$  for  $s \in (0, \tilde{T})$ , and

$$\int_0^{\tilde{T}} \mathcal{A}^* \left[ u(\tilde{X}(s)) Z(s) \right] ds = - \int_0^{\tilde{T}} p(X(s), s) Z(s) ds$$

for any  $(x, t)$  in  $D \times (0, \infty)$ . Technicalities on these derivatives can be found in [3] (Sect. 6.4). ▲

*Example 5.29* Consider the special case of the time-invariant Schrödinger equation obtained from (5.50) with  $q(x) = q$ ,  $p(x) = 0$ ,  $\alpha_i = 0$ , and  $\beta_{ij} = 2\delta_{ij}$ . The local solution for the boundary condition  $u(x) = \xi(x)$  is

$$u(x) = E^x \left[ \xi(\sqrt{2}B(T)) e^{qT} \right], \quad x \in D,$$

where  $B$  is an  $\mathbb{R}^d$ -valued Brownian motion starting at  $x \in D$  and  $T = \inf\{t > 0 : \sqrt{2}B(t) \notin D, x \in D\}$  is a stopping time. Estimates of  $u(x)$  for  $D = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 < 4\}$ ,  $q = -\sum_{i=1}^2 a_i^2$ , and  $\xi(x) = \exp(a_0 + a_1 x_1 + a_2 x_2)$  with  $a_0 = 0.5$ ,  $a_1 = 0.3$ , and  $a_2 = 0.3$  are in error relative to exact solution  $u(x) = \exp(a_0 + a_1 x_1 + a_2 x_2)$  by less than 1.5% if based on 500 independent samples of  $B$  generated at a time step  $\Delta t = 0.001$ . ◇

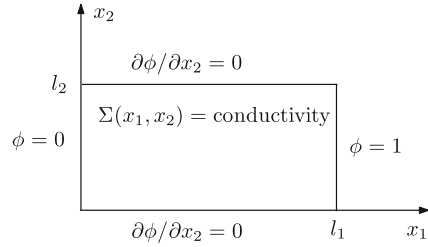
### 5.5.3.3 Mixed Boundary Value Problem

Local solutions for partial differential equations with Neumann and Dirichlet boundary conditions use diffusion processes similar to those for Dirichlet boundary value problems in the interior of their domain of definition  $D$ . However, the processes solving mixed boundary value problems need to be reflected back in  $D$  when they reach Neumann boundaries since the solution is not specified on these boundaries.

We have already applied the random walk method in Example 5.25 to solve an elementary mixed boundary value problem locally. The following example applies this method to calculate the effective conductivity for a heterogeneous material specimen.

*Example 5.30* The two-dimensional rectangular specimen in Fig. 5.2 is subjected to a voltage source with unit intensity. The potential  $\phi(x)$  in the specimen satisfies the partial differential equation

**Fig. 5.2** Two-dimensional heterogeneous specimen with unit thickness



$$\sum_{p=1}^2 \frac{\partial \Sigma(x)}{\partial x_p} \frac{\partial \phi(x)}{\partial x_p} + \Sigma(x) \Delta \phi(x) = \nabla \cdot (\Sigma(x) \nabla \phi(x)) = 0, \quad x \in D, \quad (5.65)$$

in  $D = (0, l_1) \times (0, l_2)$  with the boundary conditions  $\phi(0, x_2) = 0$  and  $\phi(l_1, x_2) = 1$  for  $x_2 \in (0, l_2)$ , and  $\partial \phi(x_1, x_2) / \partial x_2 = 0$  on  $(0, l_1) \times \{0\}$  and  $(0, l_1) \times \{l_2\}$ , where  $\Sigma(x) > 0$ ,  $x \in D$ , denotes material conductivity. It is assumed that the conductivity field  $\Sigma$  is deterministic and has continuous first order partial derivatives.

Analytical solutions for (5.65) are available in few cases. For example, if  $\Sigma(x) = \Sigma$  is a constant, that is, the material is homogeneous, the potential satisfies the differential equation  $\Delta \phi(x) = 0$ ,  $x \in D$ , so that  $\phi(x) = x_1 / l_1$ ,  $x \in D$ . The current passing through the specimen in Fig. 5.2 is

$$J_{\text{heter}} = \int_{\{x_1\} \times (0, l_2)} \Sigma(x) \frac{\partial \phi(x)}{\partial x_1} dx_2, \quad x_1 \in [0, l_1], \quad (5.66)$$

for a heterogeneous specimen with conductivity  $\Sigma(x)$  and  $J_{\text{homog}} = \Sigma l_2 / l_1$  for a homogeneous specimen with a constant conductivity  $\Sigma$ . The effective or apparent conductivity  $\Sigma_{\text{eff}}$  for the heterogeneous specimen is the conductivity of a virtual homogeneous specimen with conductivity such that  $J_{\text{heter}} = J_{\text{homog}}$ , which gives

$$\Sigma_{\text{eff}} = \frac{l_1}{l_2} \int_{\{x_1\} \times (0, l_2)} \Sigma(x) \frac{\partial \phi(x)}{\partial x_1} dx_2, \quad x_1 \in [0, l_1]. \quad (5.67)$$

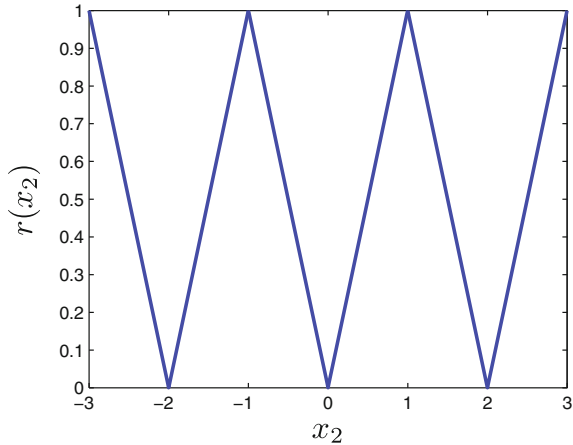
Since effective conductivity  $\Sigma_{\text{eff}}$  depends on the gradient  $\partial \phi / \partial x_1$  of  $\phi$  on a line  $\{x_1\} \times (0, l_2)$ ,  $\Sigma_{\text{eff}}$  can be calculated approximately from

$$\Sigma_{\text{eff}} \simeq \frac{l_1}{l_2} \Delta x_2 \sum_{k=0}^{n_2-1} \Sigma(l_1 - \zeta, x_{2,k}) \frac{1 - \phi(l_1 - \zeta, x_{2,k})}{\zeta}, \quad (5.68)$$

where  $x_{2,k} = (k + 1/2) \Delta x_2$ ,  $k = 0, 1, \dots, n_2 - 1$ ,  $\Delta x_2 = l_2 / n_2$ ,  $n_2 \geq 1$  is an integer, and  $0 < \zeta \ll l_1$  is a constant. The random walk method can be used to estimate  $\phi(l_1 - \zeta, x_{2,k})$ .

Let  $\tilde{X}$  be an  $\mathbb{R}^2$ -valued process with coordinates  $\tilde{X}_1 = X_1$  and  $\tilde{X}_2 = r(X_2)$ , where  $X = (X_1, X_2)$  is diffusion process defined by the stochastic differential equation

**Fig. 5.3** Function  $r(x_2)$  in the range  $[-3, 3]$  for  $l_2 = 1$



$$\begin{cases} dX_1(t) = a_1(X(t)) dt + b(X(t)) dB_1(t) \\ dX_2(t) = a_2(X(t)) dt + b(X(t)) dB_2(t), \end{cases} \quad (5.69)$$

where  $a_k(x) = \partial \Sigma(x) / \partial x_k$ ,  $k = 1, 2$ ,  $b(x) = (2 \Sigma(x))^{1/2}$ , and  $r : \mathbb{R} \rightarrow \mathbb{R}$  is a periodic function with period  $2l_2$  defined by  $r(x_2) = |x_2|$  for  $x_2 \in [-l_2, l_2]$ . Figure 5.3 shows  $r(x_2)$  in the range  $[-3, 3]$  for  $l_2 = 1$ . The generator of  $X$  coincides with that of  $\tilde{X}$  in  $D$ , and matches the differential operator in (5.65). The essential difference between  $X$  and  $\tilde{X}$  relates to their behavior at the Neumann boundaries.  $X$  can exit these boundaries while  $\tilde{X}$  is reflected back in  $D$  so that it can exist only through the Dirichlet boundaries of  $D$ . The local solution for (5.65) is

$$\phi(x) = E^x[\phi(\tilde{X}(T^*))], \quad x \in D, \quad (5.70)$$

where  $T^* = \inf\{t > 0 : \tilde{X}(t) \notin D, \tilde{X}(0) = x \in D\}$ .

Estimates for  $\Sigma_{\text{eff}}$  have been obtained from (5.68) with  $\phi(l_1 - \zeta, x_{2,k})$  calculated from (5.70) for a material specimen in  $D = (0, 10) \times (0, 4)$  with conductivity field

$$\Sigma(x) = 2 + \sin\left(\frac{q_1 \pi x_1}{l_1}\right) \sin\left(\frac{q_2 \pi x_2}{l_2}\right), \quad x \in D, \quad (5.71)$$

for  $q_1 = 3$  and  $q_2 = 2$ . The estimates can be unsatisfactory if the time step  $\Delta t$  used to generate samples of  $\tilde{X}$  is relatively large, for example,  $\Sigma_{\text{eff}} = 2.73$  for  $\zeta = 0.1$ ,  $\Delta t = 0.001$ ,  $n_2 = 5$ , and  $n = 1000$ , an error of +36% with respect to the finite element solution. The estimates of  $\Sigma_{\text{eff}}$  improve by decreasing the time step  $\Delta t$  and/or increasing  $n_2$ , for example,  $\Sigma_{\text{eff}} = 2.09$  (error +4.6%) for  $\zeta = 0.1$ ,  $\Delta t = 0.0001$ ,  $n_2 = 5$ , and  $n = 1000$ , and  $\Sigma_{\text{eff}} = 1.9619$  (error -1.91%) for  $\zeta = 0.05$ ,  $\Delta t = 0.00001$ ,  $n_2 = 20$ , and  $n = 1000$ . A useful discussion on the construction of the local solution for (5.65) when  $\Sigma(x)$  has discontinuities can be found in [8].  $\diamond$

### 5.5.4 Girsanov's Theorem

Prior to stating and proving Girsanov's theorem, we define an exponential process, give properties of this process, and discuss transformations of probability measures. The presentation is based on developments in [12] (Sects. 8.7, 8.8, 8.9).

Let  $B$  be a Brownian motion on a probability space  $(\Omega, \mathcal{F}, P)$  and let  $Y$  be a member of the space  $\mathcal{H}[0, \tau]$  of real-valued processes that are adapted to the Brownian filtration  $\mathcal{F}_t = \sigma(B(s) : 0 \leq s \leq t)$  and are  $\mathcal{B}[0, \tau] \times \mathcal{F}_\tau$ -measurable such that  $P(\int_0^\tau Y(t)^2 dt < \infty) = 1$  (Sect. 4.4).

**Definition 5.7** The process

$$Z(t) = \exp\left[\int_0^t Y(s) dB(s) - \frac{1}{2} \int_0^t Y(s)^2 ds\right], \quad t \in [0, \tau], \quad (5.72)$$

with  $Y \in \mathcal{H}[0, \tau]$  is called exponential process.

*Example 5.31* Consider the special case of (5.72) with  $h \in L^2[0, \tau]$  in place of  $Y$ . Then

$$\begin{aligned} Z(t) &= \exp[V(t)] = \frac{\exp(\int_0^t h(s) dB(s))}{E[\exp(\int_0^t h(s) dB(s))]}, \text{ where} \\ V(t) &= \int_0^t h(s) dB(s) - \frac{1}{2} \int_0^t h(s)^2 ds. \end{aligned} \quad (5.73)$$

Alternative definitions of this process are  $dZ(t) = h(t) Z(t) dB(t)$  with  $Z(0) = 1$  and  $Z(t) = 1 + \int_0^t V(s) dB(s)$ , and show that  $Z(t)$  is an  $\mathcal{F}_t$ -martingale.  $\diamond$

*Proof* The expression of  $Z(t)$  in (5.73) holds since  $\int_0^t h(s) dB(s)$  is a Gaussian variable with mean 0 and variance  $\int_0^t h(s)^2 ds$ . Itô's formula applied to the mapping  $V(t) \mapsto Z(t) = \exp(V(t))$  gives  $dZ(t) = Z(t) dV(t) + (1/2) Z(t) d[V, V](t)$  so that  $dZ(t) = h(t) Z(t) dB(t)$  since  $d[V, V](t) = h(t)^2 dt$ . The integral form of this equation is  $Z(t) - 1 = \int_0^t h(s) Z(s) dB(s)$  since  $Z(0) = 1$  so that  $Z(t)$  is an  $\mathcal{F}_t$ -martingale in  $[0, \tau]$  by Theorem 4.4.  $\blacktriangle$

**Theorem 5.20** If  $Y$  and  $Z$  are as in Definition 5.7 and  $E[Z(t)] = 1$  for all  $t \in [0, \tau]$ , then  $Z(t)$  is an  $\mathcal{F}_t$ -martingale in  $[0, \tau]$ .

*Proof* Note first that  $E[Z(t)] = E[Z(0)] = 1$  for all  $t \in [0, \tau]$  if  $Z(t)$  is the process in Example 5.31. The condition  $E[Z(t)] = 1$  is difficult to verify for  $Z(t)$  in (5.72). It is preferable to use the requirement  $E[\exp(\int_0^\tau Y(t)^2 dt/2)] < \infty$ , referred to as the Novikov condition, that is stronger than  $E[Z(t)] = 1$ ,  $t \in [0, \tau]$ .

The set function  $Q$  defined by  $dQ = Z(\tau) dP$ , or its integral form  $Q(A) = \int_A Z(\tau) dP$  is a probability measure since it is countably additive and  $Q(\Omega) = \int_\Omega Z(\tau) dP = E[Z(\tau)] = 1$  by assumption. Also, the probability measures  $Q$  and  $P$  are equivalent since  $Q(A) = 0$  if and only if  $P(A) = 0$ ,  $A \in \mathcal{F}$ .

It can be shown that  $Q_t$  defined by  $dQ_t = Z(t) dP$  is the restriction of  $Q$  to  $\mathcal{F}_t$ . For  $s \leq t$  and  $A \in \mathcal{F}_s$  arbitrary, we have  $\int_A E_P[Z(t) | \mathcal{F}_s] dP = \int_A Z(t) dP = \int_A dQ_t = Q(A)$  by properties of the conditional expectation and the definition of  $Q_t$ , which implies  $E_P[Z(t) | \mathcal{F}_s] = Z(s)$  since  $Q(A) = \int_A dQ_s = \int_A Z(s) dP$ . We have  $E_P[Z(t) | \mathcal{F}_s] = Z(s)$  for  $A \in \mathcal{F}_s$  arbitrary,  $E[Z(t)]$  exists and is finite, and  $Z(t) \in \mathcal{F}_t$ , so that  $Z(t)$  is an  $P$ -martingale in  $[0, \tau]$ .  $\blacktriangle$

**Theorem 5.21** *Let  $B(t)$ ,  $0 \leq t \leq 1$ , be a Brownian motion with respect to a probability measure  $P$  and let  $Q$  be defined by  $dQ = \exp(B(1) - 1/2) dP$ . Then  $M(t) = B(t) - t$  is a Brownian motion with respect to  $Q$  and*

$$\int_{\Omega} h(B(t) - t) dQ = \int_{\Omega} h(B(t)) dP, \quad (5.74)$$

or, equivalently  $E_Q[h(B(t) - t)] = E_P[h(B(t))]$ , where  $h$  is a function such that  $E_P[|h(B(t))|] < \infty$ .

*Proof* That  $Q$  is a probability measure results directly from its definition. The process  $M(t) = B(t) - t$  is a Brownian motion under  $Q$  since it starts at zero, has continuous samples, and has stationary Gaussian increments. That  $M(t) - M(s) \sim N(0, t - s)$ ,  $t > s$ , follows from  $E_Q[e^{iu(M(t) - M(s))}] = e^{-iu(t-s)} E_Q[e^{iu(B(t) - B(s))}]$  and

$$\begin{aligned} E_Q[e^{iu(B(t) - B(s))}] &= E_P[e^{iu(B(t) - B(s))} dQ/dP] = E_P[e^{iu(B(t) - B(s))} e^{B(1) - 1/2}] \\ &= E_P\{E_P[e^{iu(B(t) - B(s))} e^{B(1) - 1/2} | \mathcal{F}_t]\} = E_P\{e^{iu(B(t) - B(s))} E_P[e^{B(1) - 1/2} | \mathcal{F}_t]\} \\ &= E_P\{e^{iu(B(t) - B(s))} e^{B(t) - t/2}\} = e^{-t/2} E_P\{e^{(iu+1)(B(t) - B(s)) + B(s)}\} \\ &= e^{-t/2} E_P\{e^{(iu+1)(B(t) - B(s))}\} E_P\{e^{B(s)}\} = e^{(-u^2 + 2iu)(t-s)/2}, \end{aligned}$$

so that  $E_Q[e^{iu(M(t) - M(s))}] = \exp(-u^2(t-s)/2)$ , that is,  $M(t) - M(s) \sim N(0, t-s)$  under  $Q$ . These calculations have used the fact that  $\exp(B(t) - t/2)$  is an  $\mathcal{F}_t$ -martingale (Exercise 5.14). Similar arguments show that  $M(t)$  has independent increments. We conclude that  $M$  is a Brownian motion with respect to  $Q$ .

To prove (5.74), note that for an arbitrary deterministic function  $a(t)$  we have

$$\begin{aligned} E_P[h(B(t) - a(t)) e^{B(1) - 1/2}] &= E_P\{E_P[h(B(t) - a(t)) e^{B(1) - 1/2} | \mathcal{F}_t]\} \\ &= E_P\{h(B(t) - a(t)) E_P[e^{B(1) - 1/2} | \mathcal{F}_t]\} = E_P\{h(B(t) - a(t)) e^{B(t) - t/2}\} \\ &= \int_{-\infty}^{\infty} h(x - a(t)) e^{x - t/2} \phi(x/\sqrt{t})/\sqrt{t} dx = E_P\{h(B(t)) e^{(t-a(t))B(t)/t - (t-a(t))^2/(2t)}\}, \end{aligned}$$

where the latter equality follows by using the change of variables  $y = x - a(t)$ . Formula (5.74) results by setting  $a(t) = t$ .  $\blacktriangle$

**Theorem 5.22** (Girsanov's theorem) *Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $Y \in \mathcal{H}[0, \tau]$ , and  $Z(t)$  as in Theorem 5.20. Then  $W(t) = B(t) - \int_0^t Y(s) ds$  is a Brownian motion with respect to probability measure  $Q(A) = \int_A Z(\tau) dP$ ,  $A \in \mathcal{F}$ .*

*Proof* First note that the probability measures  $Q$  and  $P$  are equivalent. Itô's formula applied to  $Z(t)$  and  $W(t)Z(t)$  gives  $dZ(t) = Y(t)Z(t)dB(t)$  and  $d(W(t)Z(t)) = (1 +$

$Y(t) W(t) Z(t) dB(t)$ , so that  $W(t) Z(t) = \int_0^t (1 + Y(s) W(s)) Z(s) dB(s)$  implying that  $W(t) Z(t)$ ,  $0 \leq t \leq \tau$ , is a local  $P$ -martingale (Sect. 4.4.3). It can be shown that  $W(t) Z(t)$ ,  $0 \leq t \leq \tau$ , is a  $P$ -martingale, so that  $W(t)$ ,  $0 \leq t \leq \tau$ , is a  $Q$ -martingale (Exercise 5.16).

The integral form of Itô's formula applied to  $(W(t)^2 - t) Z(t)$  gives

$$(W(t)^2 - t) Z(t) = \int_0^t (2 W(s) + (W(s)^2 - s)) Z(s) dB(s),$$

so that  $(W(t)^2 - t) Z(t)$  is a  $P$ -martingale, which implies that  $W(t)^2 - t$  is a  $Q$ -martingale (Exercise 5.16). Since  $W(t)^2 = (W(t)^2 - t) + t$  and  $W(t)^2 - t$  is a  $Q$ -martingale, we have  $\langle W \rangle(t) = t$  a.s., so that  $W(t)$  is a Brownian motion with respect to  $Q$  (Example 5.8).  $\blacktriangle$

**Example 5.32** Let  $B(t)$  be a Brownian motion on a probability space  $(\Omega, \mathcal{F}, P)$ . The stochastic process  $W(t) = B(t) - \int_0^t h(s) ds$ ,  $h \in L^2[0, \tau]$ , is a Brownian motion with respect to the probability measure  $Q$  defined by  $dQ = \exp[\int_0^\tau h(s) dB(s) - (1/2) \int_0^\tau h(s)^2 ds] dP$ .  $\diamond$

*Proof* We have  $E[\int_0^\tau h(s)^2 ds] = \int_0^\tau h(s)^2 ds < \infty$  by assumption, so that  $h \in \mathcal{H}[0, \tau]$ . Since  $Z(t)$  in (5.73) satisfies the conditions of Girsanov's theorem,  $W(t) = B(t) - \int_0^t h(s) ds$  is a Brownian motion with respect to  $Q$ .  $\blacktriangle$

**Example 5.33** Let  $X$  be a diffusion process defined by  $dX(t) = a dt + b dB(t)$ ,  $t \in [0, \tau]$ , with  $X(0) = x$ , where  $a$  and  $b \neq 0$  are constants and  $B(t)$  is a Brownian motion on a probability space  $(\Omega, \mathcal{F}, P)$ . Then  $W(t) = B(t) + a t/b$  is a Brownian motion with respect to the probability measure  $Q$  defined by  $dQ = Z(\tau) dP$ , where  $Z(t) = \exp[-(a/b) B(t) - a^2 t/(2 b^2)]$ . The probability of  $\{X(t) \leq \xi\}$  is  $\Phi((\xi - x - a t)/(b \sqrt{t}))$ , and can be calculated under both  $P$  and  $Q$ .  $\diamond$

*Proof* Under  $P$ ,  $B(t)$  is a Brownian motion by assumption so that  $X(t) = x + a t + b B(t) \sim N(x + a t, b^2 t)$  and  $P(X(t) \leq \xi) = \Phi((\xi - x - a t)/(b \sqrt{t}))$ .

The Girsanov theorem with  $Y(t) = -a/b$  and

$$Z(t) = \exp \left[ \int_0^t Y(s) ds - (1/2) \int_0^t Y(s)^2 ds \right] = \exp \left[ - (a/b) B(t) - a^2 t/(2 b^2) \right]$$

implies that  $W(t) = B(t) - \int_0^t Y(s) ds = B(t) + a t/b$  is a Brownian motion under  $Q$ , so that  $X(t) = x + a t + b B(t) = x + b W(t) \sim N(x, b^2 t)$  under  $Q$ . Since  $dP/dQ_t = 1/Z(t) = \exp[(a/b) B(t) + a^2 t/(2 b^2)]$  and  $a t + b B(t) = b W(t)$ , we have

$$\begin{aligned} P(X(t) \leq \xi) &= E_P[1(X(t) \leq \xi)] = E_Q \left[ 1(x + b W(t) \leq \xi) \frac{dP}{dQ_t} \right] \\ &= E_Q \left[ 1(x + b W(t) \leq \xi) \exp \left( \frac{a}{b} B(t) + \frac{a^2 t}{2 b^2} \right) \right] \\ &= E_Q \left[ 1(x + b W(t) \leq \xi) \exp \left( \frac{a}{b} W(t) - \frac{a^2 t}{2 b^2} \right) \right]. \end{aligned}$$

This shows that  $P(X(t) \leq \xi)$  can be calculated from

$$\begin{aligned} P(X(t) \leq \xi) &= \int_{-\infty}^{(\xi-x)/b} \exp\left(\frac{a}{b}\sigma - \frac{a^2 t}{2b^2}\right) \phi(\sigma/\sqrt{t})/\sqrt{t} d\sigma \\ &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{(\xi-x)/b} \exp\left[-(\sigma - at/b)^2/(2t)\right] d\sigma \\ &= \Phi((\xi-x) - at)/(b\sqrt{t}) \end{aligned}$$

since  $W(t) \sim N(0, t)$  under  $Q$ .  $\blacktriangle$

*Example 5.34* Suppose  $X$  is defined by  $dX(t) = -\alpha X(t) dt + \sqrt{2\alpha} dB(t)$ , where  $X(0) = x$ ,  $\alpha > 0$ ,  $B(t)$  is a Brownian motion on a probability space  $(\Omega, \mathcal{F}, P)$ . The probability  $p_f(\tau) = P(\max_{0 \leq t \leq \tau} X(t) > x_{\text{cr}})$  can be estimated by

$$\hat{p}_{f,MC}(\tau) = \frac{1}{n} \sum_{i=1}^n 1\left(\max_{0 \leq t \leq \tau} \{x_i(t)\} > x_{\text{cr}}\right),$$

where  $\{x_i(t)\}$  are independent samples of  $X(t)$  and  $x_{\text{cr}}$  is a specified threshold. If very few samples of  $X(t)$  exceed  $x_{\text{cr}}$ , the estimate will be inefficient. It can be improved by measure change. For example, set  $Y(t) = h(t) = \gamma \sqrt{2\alpha}$  in (5.73), where  $\gamma$  is an arbitrary constant, so that  $Z(t) = \exp[\gamma \sqrt{2\alpha} B(t) - \gamma^2 \alpha t]$  and  $dQ_t = Z(t) dP$ ,  $t \in [0, \tau]$ . The process  $W(t) = B(t) - \gamma \sqrt{2\alpha} t$  is a Brownian motion in  $[0, \tau]$  under  $Q = Q_\tau$  so that  $dX(t) = (-\alpha X(t) + 2\alpha \gamma) dt + \sqrt{2\alpha} dW(t)$  is a diffusion process under  $Q$  driven by the Brownian motion  $W$ . Accordingly,  $p_f(\tau)$  can be estimated by

$$\hat{p}_{f,IMC}(\tau) = \frac{1}{n} \sum_{i=1}^n 1\left(\max_{0 \leq t \leq \tau} \{x_i(t)\} > x_{\text{cr}}\right) \left(\frac{dP}{dQ}\right)_i(\tau),$$

where  $\{x_i(t)\}$  are independent samples of  $X(t)$  under  $Q$  since

$$p_f(\tau) = E_P \left[ 1\left(\max_{0 \leq t \leq \tau} X(t) > x_{\text{cr}}\right) \right] = E_Q \left[ 1\left(\max_{0 \leq t \leq \tau} \tilde{X}(t) > x_{\text{cr}}\right) (dP/dQ)(\tau) \right].$$

The estimates  $\hat{p}_{f,MC}(\tau)$  are zero if based on  $n = 500, 1,000, 5,000$ , and  $10,000$  independent samples of  $X(t)$ , and  $0.5 \times 10^{-4}$  if based on  $n = 100,000$  independent samples of this process. Corresponding estimates  $\hat{p}_{f,IMC}(\tau)$  with  $\gamma = 4$  are  $0.2340 \times 10^{-4}$ ,  $0.3384 \times 10^{-4}$ ,  $0.3618 \times 10^{-4}$ ,  $0.3402 \times 10^{-4}$ , and  $0.3981 \times 10^{-4}$  for  $n = 500, 1,000, 5,000, 10,000$ , and  $100,000$  samples, respectively. These numerical results are for  $\alpha = 1$ ,  $x_{\text{cr}} = 2.5$ , and  $\tau = 1$ .  $\diamond$

## 5.6 Exercises

**Exercise 5.1.** Develop a recurrence formula for calculating the stochastic integral  $\int_0^t C(s-)^n dC(s)$ , where  $C$  denotes a compound Poisson process and  $n \geq 1$  is an integer.

**Exercise 5.2.** Show that the summation in (5.7) is a semimartingale.

**Exercise 5.3.** Let  $X(t) = (X_1(t) = \cos(B(t)), X_2(t) = \sin(B(t)))$  be an  $\mathbb{R}^2$ -valued stochastic process depending on a Brownian motion  $B(t)$ . Show that  $X(t)$  is a diffusion process and find the stochastic differential equation defining this process.

**Exercise 5.4.** Let  $X(t)$ ,  $t \geq 0$ , be the Ornstein–Uhlenbeck process in Example 5.10. Show that  $Y(t) = \exp(X(t))$  is a diffusion process by constructing the stochastic differential equation that defines this process.

**Exercise 5.5.** Find the relationship between the Itô and the Stratonovich integrals  $\int_0^t C(s-) dC(s)$  and  $\int_0^t C(s-) \circ dC(s)$ .

**Exercise 5.6.** Construct the Itô and the Stratonovich differential equations for  $X(t) = \exp(B(t))$ ,  $t \geq 0$ , where  $B(t)$  is a Brownian motion.

**Exercise 5.7.** Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a differentiable function and consider the Itô stochastic integral equation  $X(t) = X(0) + \int_0^t h(X(s)) dB(s) + (1/2) \int_0^t h(X(s)) h'(X(s)) ds$ . Find the solution of this equation.

**Exercise 5.8.** Show that the integration by parts formula  $\int_0^t h(s) dB(s) = h(t)B(t) - \int_0^t h'(s) B(s) ds$  holds for  $h \in C^1[0, \infty)$ .

**Exercise 5.9.** Calculate the first four moments of  $X(t)$  in Example 5.24 for increasing values of  $\lambda$  and a fixed  $\alpha > 0$  under the constraint  $\lambda E[Y_1^2] = 2\alpha$ . Comment on the dependence of the skewness and kurtosis of  $X(t)$  on  $\lambda$ .

**Exercise 5.10.** Show that the expectation of the local time  $L(t)$  in (5.46) exists and is  $E[L(t)] = \sqrt{2t/\pi}$  for all  $t \geq 0$ . Estimate  $E[L(t)]$  from samples of  $|B(t)|$ .

**Exercise 5.11.** Write a Monte Carlo algorithm for estimating the Prandtl stress function in Example 5.27.

**Exercise 5.12.** Let  $\hat{U}(x) = (1/n) \sum_{i=1}^n Z_i + (1/n) \sum_{i=1}^n I_i$  be an estimator of  $u(x)$  in (5.57) corresponding to  $n$  independent samples of  $X$ , where  $Z_i$  and  $I_i$  denote independent copies of  $\xi(X(T))$  and  $\int_0^T p(X(\sigma)) d\sigma$ , respectively. Show that the estimator  $\hat{U}(x)$  is unbiased and weakly consistent, that is,  $E[\hat{U}(x)] = u(x)$  and  $P(|\hat{U}(x) - u(x)| > \varepsilon) \rightarrow 0$  as  $n \rightarrow \infty$  for arbitrary  $\varepsilon > 0$ .

**Exercise 5.13.** Let  $R \in L^1(P)$  be non-negative and  $P$  and  $Q$  be probability measures on a measurable space  $(\Omega, \mathcal{F})$  such that  $dQ = R dP$ . For  $X \in L^1(Q)$  and a sub- $\sigma$ -field  $\mathcal{G}$  of  $\mathcal{F}$ , show that  $E_Q[X | \mathcal{G}] = E_P[X R | \mathcal{G}] / E_P[R | \mathcal{G}]$  holds  $P$ -almost surely.



*Hint* Since  $E_P[|X R|] = \int_{\Omega} |X| R dP = \int_{\Omega} |X| dQ = E_Q[|X|] < \infty$  by assumption,  $E_P[X R | \mathcal{G}]$  is defined. Properties of conditional expectation give

$$\int_G E_P[X R | \mathcal{G}] dP = \int_G X R dP = \int_G X dQ = \int_G E_Q[X | \mathcal{G}] dQ$$

and

$$\begin{aligned} \int_G E_Q[X | \mathcal{G}] dQ &= \int_G E_Q[X | \mathcal{G}] R dP = \int_G E_P\{E_Q[X | \mathcal{G}] R | \mathcal{G}\} dP \\ &= \int_G E_Q[X | \mathcal{G}] E_P[R | \mathcal{G}] dP \end{aligned}$$

for any  $G \in \mathcal{G}$ . These observations imply the stated relationship.

**Exercise 5.14.** Show that  $\exp(B(t) - t/2)$  is an  $\mathcal{F}_t$ -martingale, where  $\mathcal{F}_t = \sigma(B(s) : 0 \leq s \leq t)$  denotes the filtration generated by a Brownian motion  $B(t)$ .

**Exercise 5.15.** Apply Itô's formula to  $Z(t)$ ,  $W(t)$ , and  $(W(t)^2 - t)Z(t)$  to confirm the expressions of the differential and integral form of these processes used in the proof of Theorem 5.22.

**Exercise 5.16.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $(\mathcal{F}_t)_{t \geq 0}$  a filtration generated by a Brownian motion. For  $Y$  and  $Z$  as in Theorem 5.20, show that an  $\mathcal{F}_t$ -adapted process  $Y(t)$ ,  $0 \leq t \leq \tau$ , is a  $Q$ -martingale if and only if  $Y(t)Z(t)$  is a  $P$ -martingale, where  $Q(A) = \int_A Z(\tau) dP$ ,  $A \in \mathcal{F}$ .

*Hint* Suppose  $Y(t)Z(t)$  is a  $P$ -martingale. Use Exercise 5.13 with  $(Y(t), Z(\tau), \mathcal{F}_s)$  in place of  $(X, R, \mathcal{G})$  for  $s \leq t$ , the fact that  $Z(t)$  is a  $P$ -martingale (Theorem 5.20), and properties of conditional expectation. Show that  $E_P[Y(t)Z(\tau) | \mathcal{F}_s] = Y(s)Z(s)$ , which gives  $E_Q[Y(t) | \mathcal{F}_s] = Y(s)$ . Similar arguments can be used to show that, if  $Y(t)$  is a  $Q$ -martingale, then  $Y(t)Z(t)$  is a  $P$ -martingale.

**Exercise 5.17.** Construct a Monte Carlo algorithm for calculating the probability  $p_f(\tau) = P(\max_{0 \leq t \leq \tau} X(t) > x_{\text{cr}})$  based on Girsanov's theorem, where  $X$  is a diffusion process defined by  $dX(t) = (a_1 X(t) + a_2 X(t)^3) dt + b dB(t)$ ,  $t \in [0, \tau]$ , with  $X(0) = 0$ . Test the algorithm against direct Monte Carlo simulation for constants  $a_1$ ,  $a_2$ ,  $b$ , and  $x_{\text{cr}}$  of your choice.

## References

1. Borelli AP, Chong KP (2000) Elasticity in engineering mechanics. Wiley, New York
2. Brabenec RL (1990) Introduction to real analysis. PWS-KENT Publishing Company, Boston
3. Chung KL, Williams RJ (1990) Introduction to stochastic integration. Birkhäuser, Boston
4. Durrett R (1996) Stochastic calculus: A practical introduction. CRC Press, New York

5. Grigoriu M (1997) Local solutions of Laplace, heat, and other equations by Itô processes. *J Eng Mech ASCE* 123(8):823–829
6. Grigoriu M (1997) Solution of some elasticity problems by the random walk method. *ACTA Mechanica* 125:197–209
7. Grigoriu M (2002) *Stochastic calculus. Applications in science and engineering*. Birkhäuser, Boston
8. Grigoriu M, Papoulia KD (2005) Effective conductivity by a probability-based local method. *J Appl Phys* 98:033706 (1–10)
9. Keskin RSO, Grigoriu M (2010) A probability-based method for calculating effective diffusion coefficients of composite media. *Probab Eng Mech* 25(2):249–254
10. Kloeden PE, Platen E (1992) *Numerical solutions of stochastic differential equations*. Springer, New York
11. Krishnan V (1984) *Nonlinear filtering and smoothing: An introduction to martingales, stochastic integrals and estimation*. Wiley, New York
12. Kuo H-H (2005) *Introduction to stochastic integration*. Springer, New York
13. Mikosch T (1998) *Elementary stochastic calculus*. World Scientific, New Jersey
14. Øksendal B (1998) *Stochastic differential equations. An introduction with applications*. Springer, New York
15. Protter P (1990) *Stochastic integration and differential equations*. Springer, New York
16. Schuss Z (1980) *Theory and applications of stochastic differential equations*. Wiley, New York
17. Snyder DL (1975) *Random point processes*. Wiley, New York
18. Steele JM (2001) *Stochastic calculus and financial applications*. Springer, New York
19. Wong E, Zakai M (1965) On the convergence of ordinary integrals to stochastic integrals. *Ann Math Stat* 36:1560–1564

## Chapter 6

# Probabilistic Models

### 6.1 Introduction

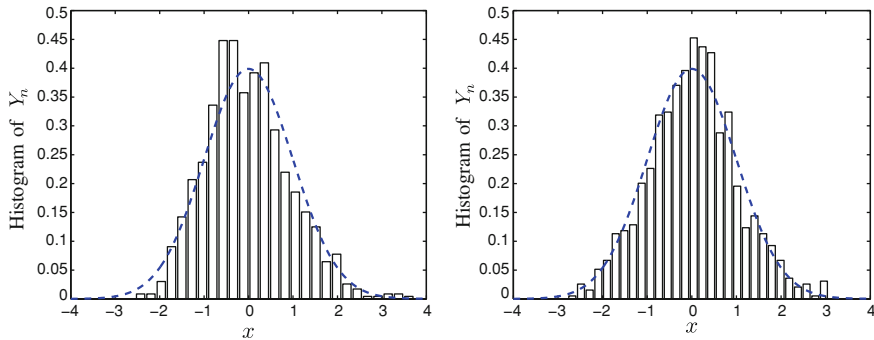
Probabilistic models are used extensively in applications to characterize properties of physical systems and describe features of the input to these systems. Generally, in applications, model construction involves two phases, referred to as model selection and model calibration.

Physics and any other available information can be used to select the functional form of the probability law of a random element up to a vector  $\theta$  of unknown or uncertain parameters that remain unspecified. For example, Beta translation fields may be used to model the spatial variation for a material conductivity since these fields take values in bounded intervals. However, the range and the shape parameters of marginal distributions for these fields as well as the parameters of their correlation functions may not be known.

Data and other information can be used to calibrate a proposed model with functional form specified up to a vector  $\theta$  of uncertain parameters. The available information can also be employed to select the optimal member of a collection of model candidates and calibrate this model [8]. Our discussion is limited to the calibration of a single proposed model. Frequentist and Bayesian methods are usually employed to characterize  $\theta$ . The frequentist approach assumes that  $\theta$  is a fixed but unknown deterministic vector and uses data to construct estimates  $\hat{\theta}$  for and confidence sets on  $\theta$ . The construction of confidence sets is usually complex when the dimension of  $\theta$  is two or larger ([27], Chap. 4). In contrast, the Bayesian method views  $\theta$  as a random element and infers its law from postulated prior densities and data.

Model construction must account for both heuristic and theoretical considerations. We give a list of items that are relevant for model construction. First, models can only provide approximate representations for physical phenomena, that may or may not be useful depending on the objective of the analysis. In short, all models are wrong but some can be useful.

Second, models provide the means for data extrapolation, which is needed since available measurements are usually insufficient for solving practical problems. For



**Fig. 6.1** Histograms of  $\sqrt{n}Y_n$  for  $X_1 \sim \text{EXP}(\lambda)$  with  $\lambda = 1$  scaled to have mean 0 and variance 1 (left panel) and  $X_1 \sim N(0, 1)$  (right panel) and  $n = 20$

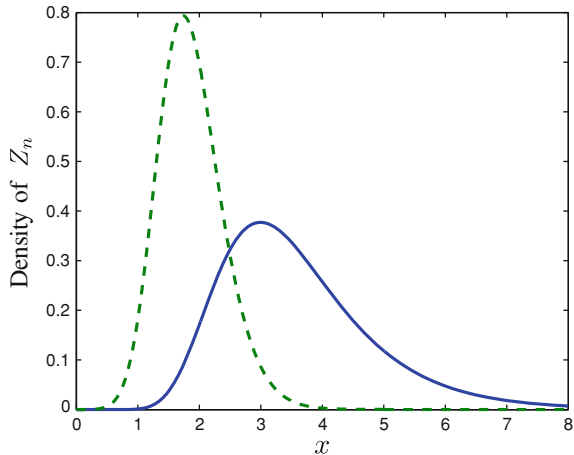
example, design of tall buildings for wind loads may require estimates of extreme wind speeds over periods exceeding 1000 years, but the length of most wind speed records is about 30–50 years. It is common to assume that yearly wind speed maxima are independent samples of Weibull or other distributions depending on uncertain parameters that can be estimated from data, and use the resulting models to estimate extreme wind speeds over 1000 years or any other period. Models would not be needed if wind records longer than 1000 years would be available.

Third, the performance of a particular model depends on its use. For example, let  $X_1, X_2, \dots$  be an iid sequence with finite variance. Suppose the mean and variance of  $X_1$  are known but the functional form of the distribution of this variable is unknown. The available information is sufficient for estimating the distribution of  $Y_n = (1/n) \sum_{i=1}^n X_i$  provided  $n$  is not very small but is inadequate to describe  $Z_n = \max_{1 \leq i \leq n} \{X_i\}$ . Figure 6.1 shows histograms of  $\sqrt{n}Y_n$ ,  $n = 20$ , for  $X_1$  assumed to be exponentially distributed with decay parameter  $\lambda = 1$ , mean 0, and variance 1 (left panel) and  $X_1 \sim N(0, 1)$  (right panel). The dotted lines in the figures are the density of  $N(0, 1)$ . Note that  $Y_n$  is approximately Gaussian irrespective of the distribution postulated for  $X_1$ . Figure 6.2 shows the densities of  $Z_n$ ,  $n = 20$ , for  $X_1$  as in Fig. 6.1. The density of  $Z_n$  depends strongly on the probability law of  $X_1$ . The available information is insufficient for estimating the probability of  $Z_n$ .

Fourth, most models are usually consistent with the physics of the problem under consideration. Physically inconsistent models have been and should be used if they provide satisfactory solutions provided they are computationally much more efficient than physically consistent models. For example, Gaussian models can be used to describe daily flows in the Nile river since they have a large mean and a small variance. However, Gaussian models are inadequate for small rivers characterized by small average daily flow and large flow variability since they may predict negative daily flows with large probabilities.

Fifth, not all uncertain parameters of a postulated model can be inferred directly from data. Properties of postulated models need to be used in this case to estimate

**Fig. 6.2** Densities of  $Z_n$  for  $X_1 \sim \text{EXP}(\lambda)$  with  $\lambda = 1$  scaled to have mean 0 and variance 1 (*dotted line*) and  $X_1 \sim N(0, 1)$  (*solid line*) for  $n = 20$



model parameters. The estimation may require additional assumptions that may further reduce model accuracy.

The following sections discuss and illustrate the construction of models for random variables and functions. Physics and data are used to select the functional form of various models and specify their uncertain parameters. Frequentist and Bayesian methods are applied for model calibration.

## 6.2 Random Variables

Let  $X$  be an  $\mathbb{R}^d$ -valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Our objective is to construct models for the probability law of  $X$  from independent samples of this random variable and any other information when available. Our construction uses Frequentist and Bayesian methods, and considers Gaussian and non-Gaussian vectors. The methods in this section are applied to construct a probabilistic model for generating directional wind speeds for hurricanes.

### 6.2.1 Gaussian Variables

Let  $X = (X_1, \dots, X_d)$  be an  $\mathbb{R}^d$ -valued Gaussian random variable with mean vector  $\mu = \{\mu_k = E[X_k], k = 1, \dots, d\}$ , covariance matrix  $\gamma = \{\gamma_{kl} = E[(X_k - \mu_k)(X_l - \mu_l)], k, l = 1, \dots, d\}$ , and density

$$f(x) = [(2\pi)^d \det(\gamma)]^{-1/2} \exp \left[ -\frac{1}{2}(x - \mu)' \gamma^{-1} (x - \mu) \right], \quad x \in \mathbb{R}^d, \quad (6.1)$$

where  $\det(\gamma)$  denotes the determinant of  $\gamma$ .

The size of the vector  $\theta$  of uncertain parameters is  $d_\theta = d + d(d + 1)/2$  if all entries of  $\mu$  and  $\gamma$  are uncertain, and  $d_\theta \leq d + d(d + 1)/2$  otherwise. Suppose that  $\mu$  and  $\gamma$  or some of their entries are uncertain and that  $n$  independent samples  $(x^{(1)}, \dots, x^{(n)})$  of  $X$  are available.

### 6.2.1.1 Frequentist Method

The method of moments and the method of likelihood function are commonly applied to construct point estimates  $\hat{\theta}$  for  $\theta$  and confidence sets ([27], Chaps. 3, 4, and 6). The method of moments calculates point estimates  $\hat{\theta}$  for  $\theta$  from equations obtained by equating  $d_\theta$  moments of  $X$  that depend on  $\theta$  to estimates of these moments obtained from data. For example, suppose  $X$  is a Gaussian random variable ( $d = 1$ ) with probability law depending on the uncertain parameters  $\theta = (\mu, \gamma)$ . In this case,

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x^{(i)} \quad \text{and} \quad \hat{\gamma} = \frac{1}{n-1} \sum_{i=1}^n (x^{(i)} - \hat{\mu})^2, \quad (6.2)$$

and probabilities of events related to  $X$  can be calculated approximately from the law of  $X$  with  $\hat{\theta} = (\hat{\mu}, \hat{\gamma})$  in place of  $\theta$ . Accordingly, the probability of  $\{X > a\}$ ,  $a \in \mathbb{R}$ , is approximated by  $P(X > a) \simeq \Phi(-(a - \hat{\mu})/\hat{\gamma}^{1/2})$ .

In the maximum likelihood method,  $\hat{\theta}$  is such that it maximizes the likelihood function

$$\ell(\theta \mid \text{data}) = \prod_{i=1}^n f(x^{(i)}; \theta), \quad (6.3)$$

viewed as function of  $\theta$ , where  $(x^{(1)}, \dots, x^{(n)})$  is the available sample and  $f(\cdot; \theta)$  denotes the density of  $X$ . We refer to  $\hat{\theta}$  as the maximum likelihood estimate (MLE) of  $\theta$ . For example, suppose  $X$  is a real-valued Gaussian variable with unknown mean  $\mu$  and variance  $\gamma$ . The likelihood function in (6.3) is

$$\ell(\theta \mid \text{data}) = \ell(\mu, \gamma \mid \text{data}) = (2\pi\gamma)^{-n/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^n \frac{(x^{(i)} - \mu)^2}{\gamma} \right]. \quad (6.4)$$

It is convenient to find the MLE  $(\hat{\mu}, \hat{\gamma})$  by maximizing the logarithm  $\ln(\ell(\theta \mid \text{data}))$  of the likelihood function. The resulting MLE estimates are  $\hat{\mu} = (1/n) \sum_{i=1}^n x^{(i)}$  and  $\hat{\gamma} = (1/n) \sum_{i=1}^n (x^{(i)} - \hat{\mu})^2$ .

Confidence sets can be used to quantify the uncertainty in the unknown parameters of a model. For example, suppose  $X$  is a real-valued Gaussian variable with unknown mean  $\mu$  and known variance  $\gamma$ . Let  $\hat{U} = (1/n) \sum_{i=1}^n X^{(i)}$  be an estimator for  $\mu$  corresponding to  $n$  independent samples of  $X$ , where  $X^{(i)}$  are independent

copies of  $X$ . Since  $\hat{U}$  is a Gaussian variable with mean  $\mu$  and variance  $\gamma/n$ ,  $\sqrt{n/\gamma}(\hat{U} - \mu) \sim N(0, 1)$  so that  $P(-a_{p/2} \leq \sqrt{n/\gamma}(\hat{U} - \mu) \leq a_{p/2}) = p$  for  $p \in (0, 1)$  and  $\Phi(a_{p/2}) = (1 - p)/2$  or, equivalently,  $P(\hat{U} - a_{p/2}\sqrt{\gamma/n} \leq \mu \leq \hat{U} + a_{p/2}\sqrt{\gamma/n}) = p$ . The interval  $I(p) = (L_1 = \hat{U} - a_{p/2}\sqrt{\gamma/n}, L_2 = \hat{U} + a_{p/2}\sqrt{\gamma/n})$  is referred to as a (symmetric) confidence set for the mean  $\mu$  of  $X$ . Note that, on average, a percentage of  $100p$  samples  $(l_1, l_2)$  of  $I(p) = (L_1, L_2)$  corresponding to sets of  $n$  independent samples of  $X$  include  $\mu$ . The confidence interval  $(L_1, L_2)$  on  $\mu$  can be used to construct a similar confidence set on  $P(X > a)$ ,  $a \in \mathbb{R}$ , since this probability is an increasing function of  $\mu$  for a fixed  $a$ . The  $p$ -confidence interval on  $P(X > a)$  is  $(L_{P,1} = (\Phi(L_1 - a)/\sigma), L_{P,2} = (\Phi(L_2 - a)/\sigma))$ .

Similar considerations can be applied in the more general case in which both parameters  $(\mu, \gamma)$  of the distribution of  $X$  are unknown. However, the construction of confidence sets is less simple ([27], Theorem 4.2.5, Sect. 4.3). Moreover, it is difficult to map these confidence sets on  $(\mu, \gamma)$  into confidence sets on the probability of events depending on  $X$ .

### 6.2.1.2 Bayesian Method

Let  $X$  be a Gaussian variable with unknown mean  $\mu$  and variance  $\gamma = 1/h$ . The uncertain parameters  $\theta = (\mu, h)$  of the law of  $X$  are viewed as random variables, so that the density of the conditional random variable  $X | \theta$  is  $f(x) = (h/(2\pi))^{1/2} \exp(-h(x - \mu)^2/2)$ . Suppose that, as previously,  $n$  independent samples  $(x^{(1)}, \dots, x^{(n)})$  of  $X$  are available and that, in addition, there is preliminary information on  $\theta$  that can be quantified by a prior density  $f'(\theta)$ . For  $X \sim N(\mu, 1/h)$ , it is common to describe the prior information on  $\theta = (\mu, h)$  by the normal-gamma density

$$f'(\theta) = f'(\mu, h) = \left(\frac{hn'}{2\pi}\right)^{1/2} \exp\left[-\frac{hn'}{2}(\mu - \mu')^2\right] \times \frac{(\zeta'v'/2)^{v'/2}}{\Gamma(v'/2)} h^{v'/2-1} \exp[-\zeta'v'h/2] \quad (6.5)$$

with parameters  $(\mu', n', \zeta', v')$ ,  $\mu' \in \mathbb{R}$ ,  $n', \zeta', v' > 0$ , that is,  $\mu | h$  is a Gaussian variable with mean  $\mu'$  and variance  $1/(hn')$  and  $h$  is a Gamma2 variable with parameters  $(\zeta', v')$ . The posterior density  $f''(\theta)$  of  $\theta$  can be calculated from

$$f''(\theta) \propto f'(\theta)\ell(\theta | \text{data}), \quad (6.6)$$

where  $\ell(\theta | \text{data})$  denotes the likelihood function and the symbol  $\propto$  means proportionality. For the prior density in (6.5),  $f''(\theta)$  is also a normal-gamma with parameters  $(\mu'', n'', \zeta'', v'')$  given by

$$\begin{aligned} n'' &= n + n', \quad v'' = n + v', \quad \mu'' = (n\bar{x} + n'\mu')/n'' \\ \zeta'' v'' &= \zeta' v' + n'(\mu')^2 + (n-1)s^2 + n\bar{x}^2 - n''(\mu'')^2, \end{aligned} \quad (6.7)$$

where  $\bar{x} = \sum_{i=1}^n x^{(i)}$  and  $s^2 = \sum_{i=1}^n (x^{(i)} - \bar{x})^2 / (n-1)$  ([31], p. 55 and Chap. 7). Prior density functions with the property that their posterior densities are of the same type are called conjugate priors. If there is no prior information on  $\theta = (\mu, h)$ , we can use the so-called noninformative or vague prior density  $f'(\mu, h) \propto 1/h$ . Under this prior information, the frequenties and the Bayesian methods use the same information.

The posterior density  $f''(\theta)$  accounts for all available information, that is, data and prior knowledge on  $\theta$ , and can be used to construct point estimates of and confidence sets on  $\theta$ . For example, the expectation of  $\theta$  with respect to the probability measure  $f''(\theta)$  can be used as an estimate for  $\theta$ . The probability of event  $\{X \in D = (a, b)\}$ ,  $a < b$ , can be calculated from

$$P(X \in D) = \int_{-\infty}^{\infty} \int_0^{\infty} \left[ \Phi((b-\mu)h^{1/2}) - \Phi((a-\mu)h^{1/2}) \right] f''(\mu, h) d\mu dh \quad (6.8)$$

since  $X | \theta$  is a Gaussian variable. Note that the uncertainty in the parameters of the law of  $X$  can be incorporated simply in the Bayesian analysis by following the procedure in (6.8), in contrast to similar calculations within the Frequentist framework that pose notable difficulties. An extensive discussion on the Bayesian method can be found in [31] (Chap. 3) and [41] (Chaps. II, III, and VIII).

## 6.2.2 Translation Variables

Let  $G = (G_1, \dots, G_d)$  be an  $\mathbb{R}^d$ -valued Gaussian variable with means  $E[G_k] = 0$ , variances  $E[G_k^2] = 1$ ,  $k = 1, \dots, d$ , and covariances  $\rho = \{\rho_{k,l} = E[G_k G_l], k, l = 1, \dots, d\}$ . Define an  $\mathbb{R}^d$ -valued non-Gaussian variable  $X = (X_1, \dots, X_d)$ , referred to as translation vector, by

$$X_k = F_k^{-1} \circ \Phi(G_k), \quad k = 1, \dots, d, \quad (6.9)$$

where  $F_k$ ,  $i = 1, \dots, d$ , are arbitrary continuous distributions (Sect. 3.7.2).

Consider an  $\mathbb{R}^d$ -valued non-Gaussian translation variable  $X$ . Suppose that (1) the functional forms of its marginal distributions  $\{F_k\}$  are known but all or some of their parameters are uncertain and collect them in a vector  $\theta^{(k)}$  and (2) the information on  $X$  consists of  $n$  independent samples  $(x^{(1)}, \dots, x^{(n)})$  and possibly prior information on the uncertain parameters in the probability law of  $X$ .

If the marginal distributions  $\{F_k\}$  of  $X$  are known, the observations  $(x^{(1)}, \dots, x^{(n)})$  can be mapped into their Gaussian images  $(g^{(1)}, \dots, g^{(n)})$  by the transformations  $g_k^{(i)} = \Phi^{-1} \circ F_k(x_k^{(i)})$  for  $k = 1, \dots, d$  and  $i = 1, \dots, n$ . Estimation procedures presented in a previous section for Gaussian variable can be applied to estimate



the covariance matrix  $\rho$  of the Gaussian image  $G$  of  $X$ . The estimate of  $\rho$  and the distributions  $\{F_k\}$  define the probability law of  $X$ .

We consider the case in which the functional forms of the distributions  $\{F_k\}$  are known but their parameters and the covariance matrix  $\rho$  are unknown. A two-dimensional translation vector  $X$  with coordinates  $X_1 \sim N(\mu, \gamma = \sigma^2)$  and  $X_2 \sim EXP(\lambda)$  depending on the unknown parameters  $\mu, \sigma$ , and  $\lambda$  is used for illustration. Note that it is not possible to calculate the Gaussian image  $(g^{(1)}, \dots, g^{(n)})$  of  $(x^{(1)}, \dots, x^{(n)})$  in this case since the mappings  $x_1^{(i)} \mapsto g_1^{(i)} = (x_1^{(i)} - \mu)/\sigma$  and  $x_2^{(i)} \mapsto g_2^{(i)} = \Phi^{-1}(1 - \exp(-\lambda x_2^{(i)}))$ ,  $i = 1, \dots, n$ , are uncertain.

### 6.2.2.1 Frequentist Method

The methods of moments can be used to calculate point estimates  $\{\hat{\theta}^{(k)}\}$  for the uncertain parameters of  $\{\theta^{(k)}\}$  from the available data  $(x^{(1)}, \dots, x^{(n)})$  and construct point estimates  $\hat{\rho}$  for the covariance matrix  $\rho$  of  $G$  from the Gaussian image  $(g^{(1)}, \dots, g^{(n)})$  of  $(x^{(1)}, \dots, x^{(n)})$  obtained via the mapping  $X \mapsto G$  with  $\{\hat{\theta}^{(k)}\}$  in place of  $\{\theta^{(k)}\}$ . For example, the approximate mappings for the two-dimensional translation vector  $X$  with coordinates  $X_1 \sim N(\mu, \sigma^2)$  and  $X_2 \sim EXP(\lambda)$  are  $x_1^{(i)} \mapsto g_1^{(i)} \simeq (x_1^{(i)} - \hat{\mu})/\hat{\sigma}$  and  $x_2^{(i)} \mapsto g_2^{(i)} \simeq \Phi^{-1}(1 - \exp(-\hat{\lambda} x_2^{(i)}))$ , where  $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$  are point estimates for  $(\mu, \sigma, \lambda)$  obtained by the method of moments. A point estimate  $\hat{\rho}$  for  $\rho$  can be obtained by the method of moments from the Gaussian images  $g^{(i)} = (g_1^{(i)}, g_2^{(i)})$  of  $x^{(i)}$ ,  $i = 1, \dots, n$ . The method is simple but its accuracy depends essentially on the sample size  $n$ . If  $n$  is sufficiently large such that the uncertainty in the point estimates  $\hat{\mu}, \hat{\sigma}, \hat{\lambda}$ , and  $\hat{\rho}$  is small, then the method can be adequate.

The method of the likelihood function can be used to construct simultaneous point estimates for both the uncertain parameters  $\{\theta^{(k)}\}$  and  $\rho$ . The MLEs for these parameters maximize the likelihood function

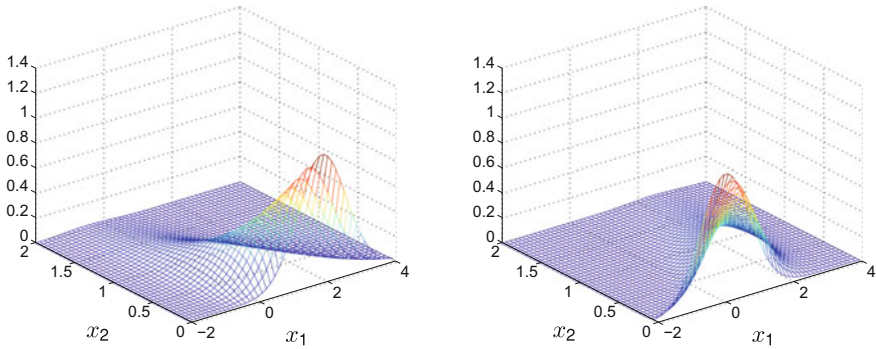
$$\begin{aligned} & \ell(\theta^{(1)}, \dots, \theta^{(d)}, \rho \mid \text{data}) \\ & \propto \det(\rho)^{-n/2} \prod_{i=1}^n \prod_{k=1}^d \frac{f_k(x_k^{(i)})}{\phi(g_k^{(i)})} \exp\left(-\frac{1}{2}(g^{(i)})' \rho^{-1} g^{(i)}\right), \end{aligned} \quad (6.10)$$

where  $g^{(i)} = (g_1^{(i)}, \dots, g_d^{(i)})$  and  $g_k^{(i)} = \Phi^{-1} \circ F_k(x_k^{(i)})$ ,  $k = 1, \dots, d$ ,  $i = 1, \dots, n$ . Analytical expressions are possible for these estimates only in special cases. Generally, optimization algorithms need to be used to find MLEs for  $\{\theta^{(k)}\}$  and  $\rho$ .

For example, the joint density of a bivariate translation vector  $X$  with coordinates  $X_1 \sim N(\mu, \sigma^2)$  and  $X_2 \sim EXP(\lambda)$  is

$$f(x_1, x_2) = \frac{\lambda}{\sigma \sqrt{2\pi(1 - \rho^2)}} \exp\left[-\lambda x_2 + \frac{g_2^2}{2} - \frac{g_1^2 - 2\rho g_1 g_2 + g_2^2}{2(1 - \rho^2)}\right], \quad (6.11)$$

where  $g_1 = (x_1 - \mu)/\sigma$  and  $g_2 = \Phi^{-1}(1 - e^{-\lambda x_2})$ . Figure 6.3 shows the density  $f(x_1, x_2)$  for  $\mu = 1$ ,  $\sigma = 1$ ,  $\lambda = 2$ , and two values of the correlation coefficient  $\rho$  of



**Fig. 6.3** Joint density of  $X$  for  $\mu = 1$ ,  $\sigma = 1$ ,  $\lambda = 2$ ,  $\rho = -0.7$  (left panel), and  $\rho = 0.7$  (right panel)

its Gaussian image,  $\rho = -0.7$  (left panel) and  $\rho = 0.7$  (right panel). The likelihood function corresponding to  $n$  independent samples  $(x^{(1)}, \dots, x^{(n)})$  of  $X$  is

$$\begin{aligned} \ell(\mu, \sigma, \lambda, \rho \mid \text{data}) &= \left( \frac{\lambda}{\sigma \sqrt{2\pi(1-\rho^2)}} \right)^n \\ &\times \exp \left[ \sum_{i=1}^n \left( -\lambda x_2^{(i)} + \frac{(g_2^{(i)})^2}{2} - \frac{(g_1^{(i)})^2 - 2\rho g_1^{(i)} g_2^{(i)} + (g_2^{(i)})^2}{2(1-\rho^2)} \right) \right], \end{aligned} \quad (6.12)$$

where  $g_1^{(i)} = (x_1^{(i)} - \mu)/\sigma$ ,  $g_2^{(i)} = \Phi^{-1}(1 - \exp(-\lambda x_2^{(i)}))$ ,  $\mu \in (-\infty, \infty)$ ,  $\sigma, \lambda > 0$ , and  $\rho \in [-1, 1]$ . The maximum likelihood estimate of  $(\mu, \sigma, \lambda, \rho)$  has to be obtained by optimization. Classical algorithms may fail since they could get trapped into local maxima. Genetic or any other global optimization algorithms usually provide accurate and reliable solutions.

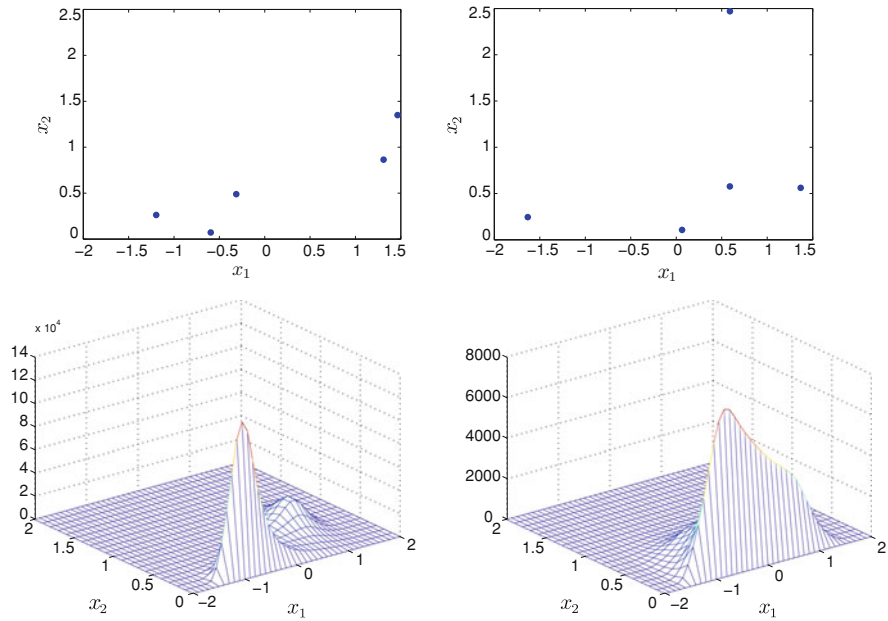
### 6.2.2.2 Bayesian Method

Suppose as previously that  $X$  is an  $\mathbb{R}^d$ -valued translation vector defined by (6.9) whose probability law depends on the uncertain parameters  $\{\theta^{(k)}\}$ ,  $k = 1, \dots, d$ , and  $\rho$ . The likelihood function corresponding to  $n$  independent samples  $(x^{(1)}, \dots, x^{(n)})$  of  $X$  is given by (6.10). The posterior density of the uncertain parameters of the probability law of  $X$  is given by

$$f''(\theta^{(1)}, \dots, \theta^{(d)}, \rho) \propto f'(\theta^{(1)}, \dots, \theta^{(d)}, \rho) \ell(\theta^{(1)}, \dots, \theta^{(d)}, \rho \mid x^{(1)}, \dots, x^{(n)}), \quad (6.13)$$

where  $f'(\theta^{(1)}, \dots, \theta^{(d)}, \rho)$  denotes the prior density of these parameters.

The posterior density in (6.13) can be used to find point estimates of and confidence sets on the uncertain parameters  $(\theta^{(1)}, \dots, \theta^{(d)}, \rho)$  and calculate, for example, the expectation  $E[1(X \in D)]$ ,  $D \subset \mathbb{R}^d$ , with respect to  $f''(\theta^{(1)}, \dots, \theta^{(d)}, \rho)$ .



**Fig. 6.4** Two data sets of 5 samples each (*top panels*) and corresponding posterior densities of  $X$  (*bottom panels*)

Let  $X$  be a bivariate translation vector with coordinates  $X_1 \sim N(\mu, \sigma^2)$  and  $X_2 \sim EXP(\lambda)$ . Assume  $\mu = 0$ , so that  $(\sigma, \lambda, \rho)$  are the uncertain parameters in the probability law of  $X$ . Suppose the data used to calculate the likelihood function in (6.12) is available and that the prior information on the uncertain parameters can be quantified by

$$f'(\sigma, \lambda, \rho) \propto 1(\sigma \in (\sigma_1, \sigma_2))1(\lambda \in (\lambda_1, \lambda_2))1(\rho \in (\rho_1, \rho_2)), \quad (6.14)$$

where  $(\sigma_1, \sigma_2) = (0.5, 1.5)$ ,  $(\lambda_1, \lambda_2) = (1, 3)$ , and  $(\rho_1, \rho_2) = (0.1, 0.9)$ . The top two panels of Fig. 6.4 show two data sets each consisting of five independent samples of  $X$  with  $(\sigma = 1, \lambda = 2, \rho = 0.7)$ . The bottom panels in the figure show posterior density functions of  $X$  corresponding to these data sets and the prior density  $f'(\sigma, \lambda, \rho)$  in (6.14). Since  $f'(\sigma, \lambda, \rho)$  is constant over its support  $(\sigma_1, \sigma_2) \times (\lambda_1, \lambda_2) \times (\rho_1, \rho_2)$ , the shapes of the posterior densities of  $(\sigma, \lambda, \rho)$  and  $X$  are determined by data. The significant differences between the densities of  $X$  in the left and the right panels of Fig. 6.4 suggest that unsatisfactory approximations may result for the posterior density of  $X$  under noninformative prior if the sample size is small.

### 6.2.3 Bounded Variables

Let  $X$  be an  $\mathbb{R}^d$ -valued random variable taking values in a bounded rectangle  $D = \times_{i=1}^d [a_i, b_i] \subset \mathbb{R}^d$ . Suppose the functional form of the distribution  $F$  of  $X$  including the parameters specifying the shape of this distributions are known. The only unknown parameters of  $F$  are those specifying its support. For example, we may take  $X_i \stackrel{d}{=} a_i + (b_i - a_i)Y_i$ ,  $i = 1, \dots, d$ , where  $Y = (Y_1, \dots, Y_d)$  is an  $\mathbb{R}^d$ -valued random variable with support  $[0, 1]^d$  and known probability law. However, the parameters  $\{a_i, b_i\}$  giving the range of  $X$  are not known. Let  $\tilde{X}$  with coordinates  $\tilde{X}_i \stackrel{d}{=} \tilde{a}_i + (\tilde{b}_i - \tilde{a}_i)Y_i$ ,  $i = 1, \dots, d$ , where  $\{\tilde{a}_i, \tilde{b}_i\}$  are estimates of  $\{a_i, b_i\}$ . Generally, the support  $\tilde{D} = \times_{i=1}^d [\tilde{a}_i, \tilde{b}_i]$  of  $\tilde{X}$  differs from  $D$ .

*Example 6.1* Let  $X$  be a Beta random variable with range  $[a, b]$  and shape parameters  $(p, q)$ ,  $p, q > 1$ , so that  $Y$  defined by  $X \stackrel{d}{=} a + (b - a)Y$  is a standard Beta variable with range  $[0, 1]$ , density  $f(y) = y^{p-1}(1-y)^{q-1}/B(p, q)$ ,  $y \in [0, 1]$ , and shape parameters  $(p, q)$ . The distribution of  $X$  is

$$F(x) = P(X \leq x) = p\left(Y \leq \frac{x-a}{b-a}\right) = \frac{1}{B(p, q)} \int_0^{\xi} y^{p-1}(1-y)^{q-1} dy = I(\xi; p, q),$$

where  $\xi = (x - a)/(b - a)$  and  $I(\xi; p, q)$  denotes the incomplete Beta function ratio.

Suppose  $(p, q)$  are known but  $D = [a, b]$  is unknown, and let  $\tilde{D} = [\tilde{a}, \tilde{b}]$  be an approximation for  $D$ . The distribution of  $\tilde{X} \stackrel{d}{=} \tilde{a} + (\tilde{b} - \tilde{a})Y$  is

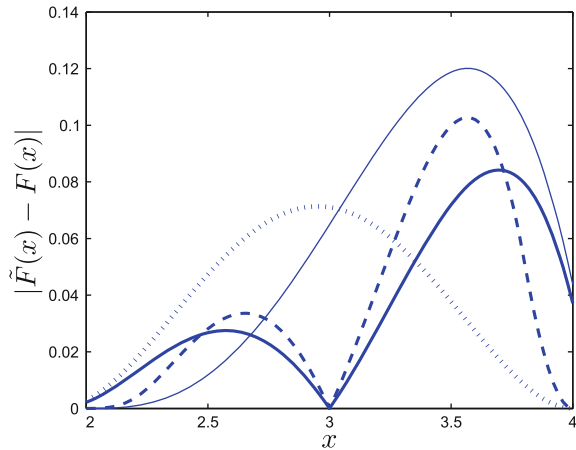
$$\tilde{F}(x) = P(\tilde{X} \leq x) = P\left(Y \leq \frac{x-\tilde{a}}{\tilde{b}-\tilde{a}}\right) = I(\tilde{\xi}; p, q),$$

where  $\tilde{\xi} = (x - \tilde{a})/(\tilde{b} - \tilde{a})$ . The difference between  $F(x)$  and  $\tilde{F}(x)$  results by direct calculations and is

$$|\tilde{F}(x) - F(x)| = \frac{1}{B(p, q)} \int_{\xi \wedge \tilde{\xi}}^{\xi \vee \tilde{\xi}} y^{p-1}(1-y)^{q-1} dy.$$

Figure 6.5 shows errors  $|\tilde{F}(x) - F(x)|$  in the range  $[2, 4]$  for  $p = 3$ ,  $q = 2$ ,  $a = 2$ ,  $b = 4$ ,  $\{\tilde{a}, \tilde{b}\} = \{1.8, 4.2\}$  (heavy solid line),  $\{2.8, 3.8\}$  (dashed line),  $\{1.8, 4.0\}$  (dotted line), and  $\{2.0, 4.2\}$  (thin solid line). The error depends strongly on differences between  $D$  and  $\tilde{D}$ . Moreover, the image  $\tilde{G} = \Phi^{-1} \circ \tilde{F}(X)$  of  $X$  based on the approximate range of this random variable is not Gaussian since  $P(\Phi^{-1} \circ \tilde{F}(X) \leq x) = P(X \leq \tilde{F}^{-1} \circ \Phi(x)) = F(\tilde{F}^{-1} \circ \Phi(x)) \neq \Phi(x)$ .  $\diamond$

**Fig. 6.5**  $|\tilde{F}(x) - F(x)|$  for  $p = 3, q = 2, a = 2, b = 4$ ,  $\{\tilde{a}, \tilde{b}\} = \{1.8, 4.2\}$  (heavy solid line),  $\{2.8, 3.8\}$  (dashed line),  $\{1.8, 4.0\}$  (dotted line), and  $\{2.0, 4.2\}$  (thin solid line)



### 6.2.4 Directional Wind Speed for Hurricanes

Hurricanes occur at random times on the United States Gulf and East coasts, and are characterized by high wind speeds with random directions. Model construction for hurricane is largely based on data since hurricane physics is rather complex.

It has been proposed in [14] to model (1) the arrival times  $T_1 < T_2 < \dots$  of hurricanes at a site during the hurricane season by a homogeneous Poisson process  $N(t)$  with intensity  $\lambda > 0$  and (2) the wind speed during distinct hurricanes by iid  $\mathbb{R}^d$ -valued random variables giving wind speeds in  $d \geq 1$  directions. Let  $F(\cdot; \theta)$  denote the distribution of the wind speed vector, where  $\theta$  collects the uncertain parameters in the expression of this distribution. The Poisson process  $N(t)$  for hurricane arrivals and the distribution  $F(\cdot; \theta)$  specify the functional form of the probability law of the hurricane process. Wind speeds recorded during hurricanes and the number of hurricanes over a reference period can be used to calibrate this process.

Let  $V$  be an  $\mathbb{R}^d$ -valued random variable denoting wind speeds during a hurricane. The coordinates  $\{V_k\}$  of  $V$  are wind speeds in  $d$  equally spaced directions. It is assumed that  $V$  is a translation vector with coordinates

$$V_k = F_k^{-1}(\Phi(G_k)), \quad k = 1, \dots, d, \quad (6.15)$$

where  $\{F_k\}$  denote wind speed distributions and  $\{G_k\}$ ,  $k = 1, \dots, d$ , are correlated standard Gaussian variables with covariance matrix  $\rho = \{\rho_{k,l} = E[G_k G_l]\}$ ,  $k, l = 1, \dots, d$ . Under this model, the functional form of the law of the hurricane process is completely specified by the intensity  $\lambda$  of the Poisson process  $N(t)$ , the marginal distributions  $\{F_k\}$  of  $V$ , and the covariance matrix  $\rho$  of  $G = \{G_k\}$ .

Suppose extreme wind speeds have been recorded in  $d$  directions during  $n$  distinct events occurring over a period of  $n_y$  years. This data set is used to calibrate the hurricane process. The average number  $\lambda$  of hurricane per year can be estimated by the ratio  $\hat{\lambda} = n/n_y$ . This estimate is accurate if  $n$  is relatively large.

The calibration of the probability law of  $V$  involves two steps. First, the parameters of the distributions  $\{F_k\}$  of the directional wind speeds  $\{V_k\}$  need to be estimated. Second, the covariance matrix  $\rho$  of the Gaussian image  $G$  of  $V$  has to be identified. We proceed with the first step. Since some directional wind speeds are zero, the marginal distributions  $\{F_k\}$  are modeled by

$$F_k(x) = q_k 1(x \geq 0) + (1 - q_k) \tilde{F}_k(x), \quad k = 1, \dots, d, \quad (6.16)$$

where  $q_k = P(V_k = 0)$  and  $\tilde{F}_k$  denotes the distribution of non-zero values of  $V_k$ . The functional forms of the distributions  $\{F_k\}$  are assumed to be known. The probability  $q_k$  can be approximated by the ratio  $\hat{q}_k = n_k/n$  for sufficiently large  $n_k$  and  $n$ , where  $n_k$  denotes the number of observed zero wind speeds in direction  $k = 1, \dots, d$ . The parameters of the distributions  $\{\tilde{F}_k\}$  can be estimated from the non-zero readings of directional wind speeds  $\{V_k\}$ . For example, suppose  $\{\tilde{F}_k\}$  are reverse Weibull distributions with parameters  $(\alpha_k, \eta_k, c_k)$ , a common assumption in wind studies [37]. Let  $Y$  be a Weibull random variable with parameters  $\alpha > 0$ ,  $\xi \in \mathbb{R}$ , and  $c > 0$ , and distribution

$$F(y) = \begin{cases} 1 - \exp\left[-\left(\frac{y-\xi}{\alpha}\right)^c\right], & y > \xi \\ 0 & y \leq \xi, \end{cases} \quad (6.17)$$

so that the reverse Weibull variable  $X \stackrel{d}{=} -Y$  has the distribution

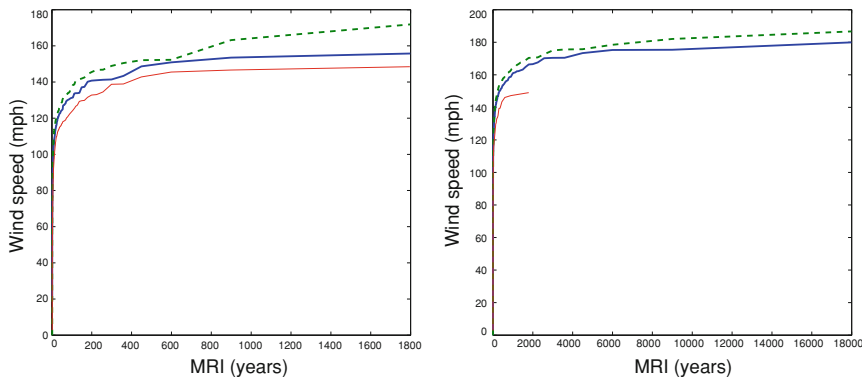
$$F(x) = P(X \leq x) = P(Y > -x) = \exp\left[-\left(\frac{\eta - x}{\alpha}\right)^c\right], \quad x \leq \eta = -\xi. \quad (6.18)$$

Moments of any order of  $Y$  can be obtained from moments  $E[\tilde{Y}^q] = \Gamma(1 + q/c)$  of the scaled random variable  $\tilde{Y}$  defined by  $Y = \xi + \alpha\tilde{Y}$  ([24], Chap. 20). For example, the mean  $\mu_y$ , variance  $\sigma_y^2$ , and skewness  $\gamma_{y,3}$  of  $Y$  are

$$\begin{aligned} \mu_y &= \xi + \alpha \Gamma(1 + 1/c) \\ \sigma_y^2 &= \alpha^2 (\Gamma(1 + 2/c) - \Gamma(1 + 1/c)^2) \\ \gamma_{y,3} &= \frac{\Gamma(1 + 3/c) - 3\Gamma(1 + 1/c)\Gamma(1 + 2/c) + 2\Gamma(1 + 1/c)^3}{(\Gamma(1 + 2/c) - \Gamma(1 + 1/c)^2)^{3/2}}. \end{aligned} \quad (6.19)$$

Let  $v_{k,1}, v_{k,2}, \dots, v_{k,n-n_k}$  be non-zero wind speed data in a direction  $k = 1, \dots, d$  recorded at a site. The method of moments, the method of maximum likelihood, the method of probability-weighted moments, and other methods can be used to estimate the parameters of this distribution ([26], Chap. 22). Extensive numerical studies in [28] suggest that the method of moments delivers satisfactory estimates for the parameters of  $\tilde{F}_k$  and is superior to, for example, the maximum likelihood method.

The covariance matrix  $\rho$  can be estimated from the Gaussian image of the available directional wind speed data defined by the mapping  $G_k = \Phi^{-1}(F_k(V_k))$ ,



**Fig. 6.6** Data and model-based wind speeds in Miami for MRIs up to 1,800 years (*left panel*) and 18,000 years (*right panel*)

$k = 1, \dots, d$ , in (6.15). This mapping cannot be used directly since the distributions  $\{F_k(x)\}$  have discontinuities at  $x = 0$ . To overcome this difficulty, all zero readings in the data set have been changed into independent samples of  $U(0, \varepsilon)$ , where  $\varepsilon > 0$  is a parameter much smaller than all non-zero wind speed readings. The distribution of the corresponding wind speed vector  $V^*$  becomes

$$F_k^*(x) = q_k \left[ \frac{x}{\varepsilon} 1(0 \leq x \leq \varepsilon) + 1(x > \varepsilon) \right] + (1 - q_k) \tilde{F}_k(x), \quad k = 1, \dots, d, \quad (6.20)$$

so that its Gaussian image is  $G_k^* = \Phi^{-1} \circ F_k^*(V_k^*)$ . This relationship maps wind data into Gaussian data, and is used to estimate the covariance matrix  $\rho^*$  of  $G^* = (G_1^*, \dots, G_d^*)$  [14].

The resulting model for the hurricane process can be used to generate directional wind speed sequences with independent and dependent coordinates following the same marginal distributions. The dependence between directional wind speeds corresponds to that between the coordinates of the Gaussian vector  $G^*$ . If the coordinates of  $G^*$  are assumed to be independent, so are the corresponding directional wind speeds.

Let  $F_{\max}$  denote the distribution of the largest wind speed  $V_{\max}$  irrespective of direction during a hurricane. The wind speed  $V_{\max}$  exceeds  $v_R$  on average every  $R$  years, where  $v_R$  is the solution of  $R = 1/[\lambda(1 - F_{\max}(v_R))]$ . The average time  $R$  is referred to as mean recurrence interval (MRI) for wind speed  $v_R$ . Figure 6.6 shows wind speeds  $v_R$  in Miami for MRIs up to 1,800 years (*left panel*) and 18,000 years (*right panel*). The heavy solid and dotted lines are predictions of the hurricane models with dependent and independent directional wind speeds described in this section. The models have been calibrated to data generated by a Monte Carlo algorithm based on physical models for hurricane wind flow ([36], Chap. 3). The thin solid lines are estimates of  $v_R$  obtained directly from this data set. Note that wind speeds for large MRIs cannot be obtained directly from data, the assumption of independence between directional wind speeds is conservative, and the hurricane model can be used to generate wind speeds of arbitrary MRI.

### 6.3 Random Functions

Let  $X(s)$ ,  $s \in D$ , be an  $\mathbb{R}^d$ -valued random function on a probability space  $(\Omega, \mathcal{F}, P)$ , where  $D$  is a subset of  $\mathbb{R}^{d'}$ . Let  $(x^{(1)}(s), \dots, x^{(n)}(s))$ ,  $s \in D$ , be  $n$  independent samples of  $X$ . Suppose the functional form of the finite dimensional distributions of  $X$  has been selected up to a vector  $\theta$  of uncertain parameters. Our objective is to estimate  $\theta$  from data and any other available information.

Estimates developed in the previous section for random variables can be extended to random functions. First,  $X(s)$ ,  $s \in D$ , is replaced with a random vector  $\tilde{X}$  consisting of values of this random function at a finite number of arguments  $s_k \in D$ . The functional form of the distribution of  $\tilde{X}$  results from the finite dimensional distributions of  $X(s)$ . Second, developments in the previous section can be applied to characterize the uncertain parameters  $\theta$  in the distribution of  $\tilde{X}$  from samples of this vector that can be extracted from samples of  $X(s)$ .

The selection of the sampling points  $s_k \in D$  used to define  $\tilde{X}$  needs to account for the temporal/spatial correlation of  $X$  which is not known. Closely spaced arguments  $\{s_k\}$  relative to the correlation distance of  $X(s)$  is inefficient and can cause numerical difficulties since it results in large vectors  $\tilde{X}$  that may have strongly dependent coordinates. On the other hand, widely spaced arguments  $\{s_k\}$  may provide an inaccurate description of the frequency content of  $X(s)$ .

*Example 6.2* Let  $X(t)$ ,  $t \geq 0$ , be a real-valued stationary Gaussian process with mean 0 and covariance function  $c(\tau) = E[X(t)X(t + \tau)] = \exp(-\lambda|\tau|)$ . Suppose that  $\lambda > 0$  is unknown and that  $n$  independent samples of  $X(t)$  have been recorded in a time interval  $[0, \bar{t}]$  at times  $0 = t_0 < t_1 < \dots < t_m = \bar{t}$ . The recorded values of  $X(t)$  constitute  $n$  independent samples of an  $\mathbb{R}^{m+1}$ -valued random variable  $\tilde{X} \sim N(0, \gamma)$ , where  $\gamma = \{c(t_i - t_j), i, j = 0, 1, \dots, m\}$ . The likelihood function of  $\lambda$  corresponding to the samples of  $\tilde{X}$  is

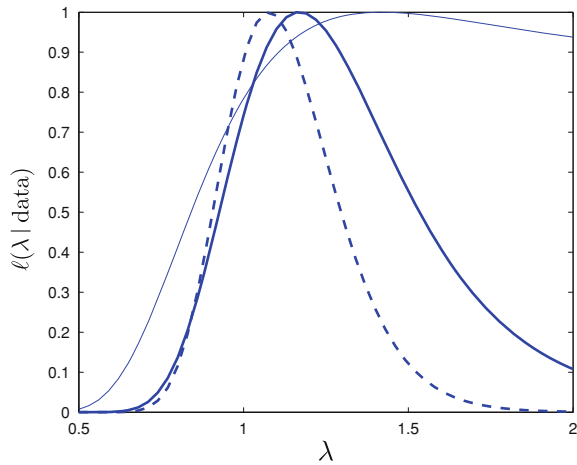
$$\ell(\lambda \mid \text{data}) \propto (\det \gamma)^{-n/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^n (\tilde{x}^{(i)})' \gamma^{-1} \tilde{x}^{(i)} \right],$$

where  $\tilde{x}^{(i)} = (x_i(t_0), x_i(t_1), \dots, x_i(t_m))'$  and  $x_i(t)$  denotes sample  $i = 1, \dots, n$  of  $X(t)$  in  $[0, \bar{t}]$ . The maximum likelihood estimate  $\hat{\lambda}$  of  $\lambda$  maximizes  $\ell(\lambda \mid \text{data})$ . The likelihood function  $\ell(\lambda \mid \text{data})$  and prior densities  $f'(\lambda)$  of  $\lambda$  can be used to calculate posterior densities  $f''(\lambda)$ .

The heavy dash, heavy solid, and thin solid lines in Fig. 6.7 are likelihood functions  $\ell(\lambda \mid \text{data})$  scaled by their largest value for  $(\bar{t} = 10, m = 100)$ ,  $(\bar{t} = 100, m = 100)$ , and  $(\bar{t} = 200, m = 100)$ , respectively, so that the corresponding sampling rates are  $\Delta t = 0.1$ ,  $\Delta t = 1$ , and  $\Delta t = 2$ . The sample of  $X(t)$  used in the analysis is for  $\lambda = 1$ . The likelihood functions for  $\Delta t > 0.1$  provide less and less information on  $\lambda$ . On the other hand, the likelihood function for  $\Delta t = 0.1$  is concentrated on the actual value of  $\lambda$  suggesting that  $\Delta t \leq 0.1$  is an adequate sampling rate for this process in the sense that it is capable of identifying the uncertain parameter  $\lambda$ .  $\diamond$



**Fig. 6.7** Scaled likelihood functions  $\ell(\lambda \mid \text{data})$  for  $(\bar{l} = 10, m = 100)$  (heavy dash line),  $(\bar{l} = 100, m = 100)$  (heavy solid line), and  $(\bar{l} = 200, m = 100)$  (thin solid line) based on a single sample of  $X(t)$  with  $\lambda = 1$



### 6.3.1 Systems with Uncertain Parameters

Suppose the probability  $P(W > w)$  provides a metric for evaluating the performance of a system with  $m$  components that have uncertain properties. Our objective is to estimate  $P(W > w)$  from the available information consisting of physics and data.

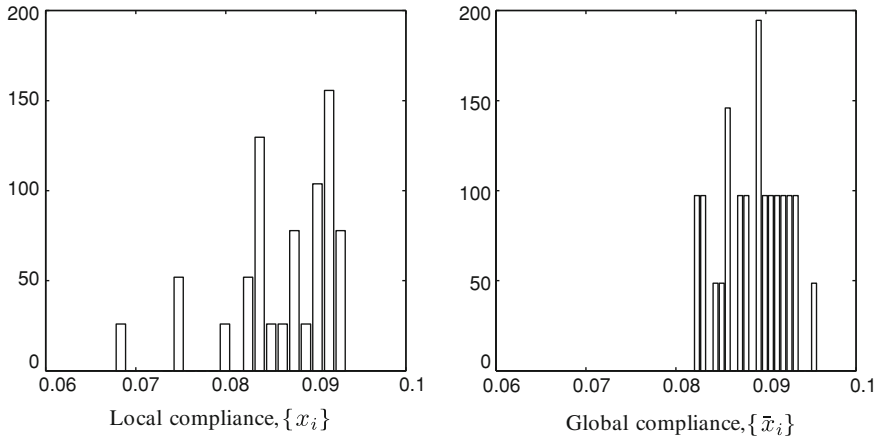
The physics implies (1)  $W = \sum_{k=1}^m c_k \bar{X}_k$ , where  $c_k$  are known constants,  $\bar{X}_k = (1/l) \int_0^l X_k(s) ds$ ,  $l > 0$ , and  $\{X_k(s)\}$  are independent copies of a real-valued random field  $X(s)$ ,  $s \geq 0$ , and (2)  $X(s)$  takes values in a bounded interval of  $(0, \infty)$ . An example of such a system is a statically determinate truss with  $m$  linear elastic components whose compliances  $X(s)$  vary randomly along them.

The data set consists of  $n$  independent samples  $(x_1, \dots, x_n)$  of  $X(s)$  at a fixed argument  $s$  and  $n$  independent samples  $(\bar{x}_1, \dots, \bar{x}_n)$  of  $\bar{X}$ . The two data sets are referred to as local and global compliances, respectively.

The available information is insufficient for constructing models for  $X(s)$  needed to calculate  $P(W > w)$ . Several assumptions are made to characterize  $X(s)$ . First, assume that  $X(s)$  is an homogeneous random field so that  $\mu = E[X(s)]$  and  $\sigma^2 = \text{Var}[X(s)] = E[(X(s) - \mu)^2]$  are space-invariant,  $c(\tau) = \sigma^2 \zeta(\tau) = E[(X(s + \tau) - \mu)(X(s) - \mu)]$  depends only on lag  $\tau$ ,  $\bar{\mu} = E[\bar{X}] = \mu$ , and

$$\bar{\sigma}^2 = \text{Var}[\bar{X}] = \left(\frac{\sigma}{l}\right)^2 \int_{-l}^l (l - |\tau|) \zeta(\tau) d\tau = \sigma^2 h(\lambda)^2. \quad (6.21)$$

Since  $|\zeta(\tau)| \leq 1$ , we have  $\bar{\sigma}^2 \leq \sigma^2$ . Equality corresponds to the case of a perfectly correlated random field  $X(s)$ , that is,  $\zeta(\tau) = 1$ . Figure 6.8 shows the available data set consisting of  $n = 30$  measurements of  $X(s)$  at an arbitrary  $s$  and  $n = 30$  measurements of  $\bar{X}$  for  $l = 20$ . The estimates of the mean and standard deviation of  $X(s)$  and  $\bar{X}$  are  $(\hat{\mu} = 0.08644, \hat{\sigma} = 0.00611)$  and  $(\hat{\bar{\mu}} = 0.08862, \hat{\bar{\sigma}} = 0.00363)$ , respectively. The approximate equality  $\hat{\mu} \simeq \hat{\bar{\mu}}$  is consistent with the



**Fig. 6.8** Data on  $X(s)$  and  $\bar{X}$

postulated stationarity of  $X(s)$  and the inequality  $\hat{\sigma} < \hat{\sigma}$  indicates that  $X(s)$  is not perfectly correlated.

Second, assume that the correlation of  $X(s)$  is exponential. We model  $X(s)$  as an exponentially correlated diffusion process, that is,

$$\begin{aligned} X(s) &= \mu + \sigma Y(s), \quad s \in [0, l] \\ dY(s) &= -\lambda Y(s) ds + b(Y(s)) dB(s), \quad \lambda > 0, \end{aligned} \quad (6.22)$$

with diffusion coefficients

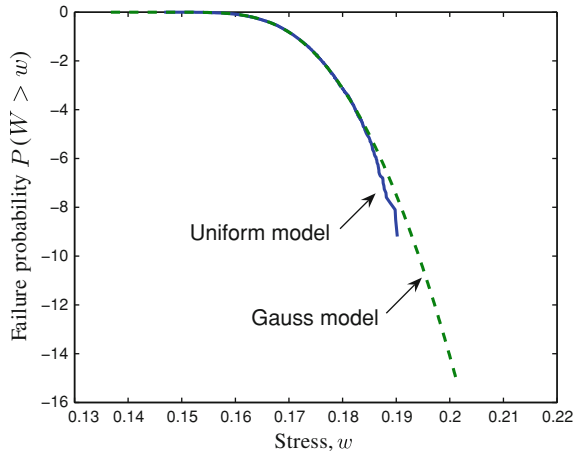
$$\begin{aligned} b(y) &= (2\lambda)^{1/2} && \text{(Gaussian distribution)} \\ b(y) &= (\lambda(3 - y^2))^{1/2}, \quad y \in [-\sqrt{3}, \sqrt{3}], && \text{(Uniform distribution),} \end{aligned} \quad (6.23)$$

where  $B$  denotes a Brownian motion. Note that the models defined by (6.22) and (6.23) depend on the uncertain parameters  $(\mu, \sigma, \lambda)$  and have the same second moment properties but different distributions. Also, the Gaussian model is physically inconsistent since the compliance field is bounded and strictly positive. This completes the first phase of model construction, that is, the selection of functional forms for the probability law of  $X(s)$ .

Model calibration, that is, the estimation of  $(\mu, \sigma, \lambda)$ , constitutes the second phase of model construction. We have found  $\hat{\mu} = 0.08644 \simeq \bar{\mu}$ ,  $\hat{\sigma} = 0.00611$ , and  $\hat{\sigma} = 0.00363$  from the available data. Since  $\zeta(\tau) = E[Y(s + \tau)Y(s)] = \exp(-\lambda|\tau|)$  for the postulated models, the function  $h(\lambda)$  in (6.21) has the expression

$$h(\lambda) = \frac{\sqrt{2(\lambda l + \exp(-\lambda l) - 1)}}{\lambda l}, \quad (6.24)$$

**Fig. 6.9** Failure probability  $P(W > w)$  under uniform and Gaussian models



which, together with the estimates of  $\bar{\sigma}$  and  $\sigma$ , gives  $\hat{\lambda} = 0.2205$ . The models for  $X(s)$  are completely specified by replacing  $(\mu, \sigma, \lambda)$  with their point estimates  $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ , so that  $P(W > w)$  can now be calculated. Alternative procedures can be used to quantify the uncertainty in  $(\mu, \sigma, \lambda)$  by, for example, the posterior density of these parameters [17].

For  $(\mu, \sigma, \lambda)$  set equal to  $(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$  the mean and variance of  $W$  are

$$W = \sum_{k=1}^m c_k \bar{X}_k \sim \left( \mu_w = \hat{\mu} \sum_{k=1}^m c_k, \sigma_w^2 = \hat{\sigma}^2 h(\hat{\lambda}) \sum_{k=1}^m c_k^2 \right),$$

so that  $P(W > w) = \Phi((\mu_w - w)/\sigma_w)$  under the Gaussian model. Since  $P(W > w)$  is not available analytically under the uniform model, it has been obtained by Monte Carlo simulation. Figure 6.9 shows the probability  $P(W > w)$  under the models in (6.22) and (6.23). Although the Gaussian model is physically inconsistent, it provides a satisfactory approximation for  $P(W > w)$ .

### 6.3.2 Inclusions in Multi-Phase Materials

The geometry and spatial distribution of inclusions can affect significantly global properties of multi-phase materials, for example, strength and permeability of concrete and various properties of other type of composite materials [5]. This section constructs and calibrates a probabilistic model capable of generating virtual inclusions with arbitrary geometry that are statistically consistent with a particular population. The presentation is based on developments in [18].

### 6.3.2.1 Spherical Harmonics

Consider an arbitrary inclusion occupying a bounded subset  $D$  of  $\mathbb{R}^3$ . Set the origin  $O$  of this space at the centroid of  $D$ , and let  $g(\theta, \varphi)$  denote the distance from  $O$  to the boundary  $\partial D$  of  $D$ , where  $(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]$  are spherical coordinates. It is assumed that  $g$  is a smooth function and  $D$  is star-like, that is, every ray from the origin of  $\mathbb{R}^3$  intersects the boundary  $\partial D$  of  $D$  at a single point [6, 29, 34]. The relationship between spherical and cartesian coordinates,  $(\theta, \varphi, g(\theta, \varphi))$  and  $(x_1, x_2, x_3)$ , is given by  $x_1 = g(\theta, \varphi) \sin(\theta) \cos(\varphi)$ ,  $x_2 = g(\theta, \varphi) \sin(\theta) \sin(\varphi)$ , and  $x_3 = g(\theta, \varphi) \cos(\theta)$ . If  $g(\theta, \varphi) = 1$  for all  $\theta \in [0, \pi]$  and  $\varphi \in [0, 2\pi]$ , then  $D$  is the unit sphere in  $\mathbb{R}^3$ .

We use spherical harmonics to represent  $g(\theta, \varphi)$  since these functions provide efficient representations for inclusions with arbitrary shapes [11, 22, 34]. The representations of  $g(\theta, \varphi)$  are truncated Fourier series in spherical harmonics of the type

$$g_n(\theta, \varphi) = \sum_{k=0}^n y_k(\theta, \varphi) = \sum_{k=0}^n \sum_{h=-k}^k d_{k,h} u_{k,h}(\theta, \varphi), \quad n = 0, 1, \dots, \quad (6.25)$$

on  $[0, \pi] \times [0, 2\pi]$ , where

$$\begin{aligned} y_k(\theta, \varphi) &= a_{k,0} p_k(\cos(\theta)) \\ &+ \sum_{h=1}^k (a_{k,h} \cos(h\varphi) + b_{k,h} \sin(h\varphi)) p_{k,h}(\cos(\theta)), \quad k = 0, 1, \dots, \end{aligned} \quad (6.26)$$

are spherical harmonic functions,

$$\begin{aligned} a_{k,0} &= d_{k,0} \left( \frac{2k+1}{4\pi} \right)^{1/2} \\ a_{k,h} &= (-1)^h \left( \frac{(2k+1)(k-h)!}{4\pi(k+h)!} \right)^{1/2} 2\Re[d_{k,h}] \\ b_{k,h} &= -(-1)^h \left( \frac{(2k+1)(k-h)!}{4\pi(k+h)!} \right)^{1/2} 2\Im[d_{k,h}], \end{aligned} \quad (6.27)$$

$$d_{k,h} = \int_{-\pi}^{\pi} \int_0^{\pi} g(\theta, \varphi) u_{k,h}^*(\theta, \varphi) \sin(\theta) d\theta d\varphi, \quad (6.28)$$

$$u_{k,h}(\theta, \varphi) = \left( \frac{(2k+1)(k-h)!}{4\pi(k+h)!} \right)^{1/2} p_{k,h}(\cos(\theta)) e^{ih\varphi} \quad (6.29)$$

$$p_{k,h}(x) = (1-x^2)^{h/2} \frac{d^h}{dx^h} p_k(x), \quad (6.30)$$

denote associated Legendre polynomials, and

$$p_k(x) = \frac{1}{2^k k!} \frac{d^k}{dx^k} (x^2 - 1)^k \quad (6.31)$$

are Legendre polynomials. If  $g$  has second order continuous partial derivatives, the sequence of function  $g_n$  in (6.25) converges absolutely and uniformly to  $g$  in  $[0, \pi] \times [0, 2\pi]$  as  $n \rightarrow \infty$ . If  $g$  is square integrable on the unit sphere, the convergence  $g_n \rightarrow g$  as  $n \rightarrow \infty$  is in the mean square sense ([32], Theorem 9, p. 726).

The first term  $a_{0,0}p_0(\cos(\theta)) = a_{0,0}$  of the series  $g_n(\theta, \varphi)$  in (6.25) gives the radius of a sphere providing a first order approximation for an inclusion. The subsequent terms in the expression of  $g_n(\theta, \varphi)$  capture differences between inclusions and their approximating spheres. Generally,  $n \simeq 20$  terms suffice to represent accurately particles with complex geometry [11]. The number of coefficients  $(a_{k,0}, a_{k,l}, b_{k,l})$  in the expression of  $g_n$  is  $(n+1)^2$ .

### 6.3.2.2 Data Set

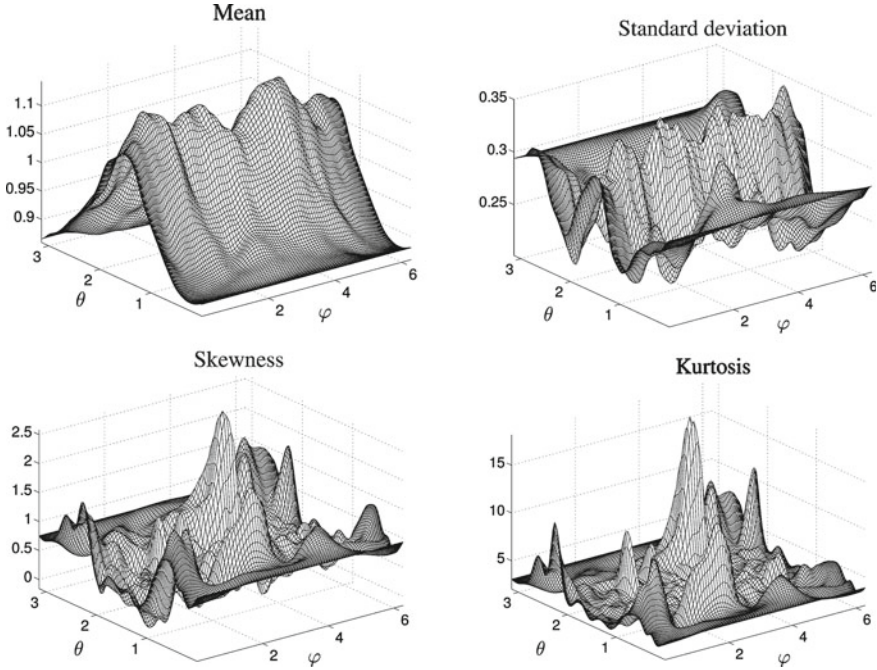
Measurements performed on  $n_s = 128$  aggregates commonly found in concrete with sizes ranging from about 4 to 28 mm have been used to calculate the coefficients  $(a_{k,0}^{(i)}, a_{k,h}^{(i)}, b_{k,h}^{(i)})$  of  $g_n^{(i)}(\theta, \varphi)$ ,  $i = 1, \dots, n_s$ , in (6.25) and (6.27). The resulting coefficients have been organized in a matrix  $a_{\text{coeff}}$  with  $n_s = 128$  columns corresponding to the collection of aggregates and  $31^2 = 961$  rows for  $n = 30$ . For example, column  $i$  in  $a_{\text{coeff}}$  consists of the coefficients  $(a_{k,0}^{(i)}, a_{k,h}^{(i)}, b_{k,h}^{(i)})$  for  $k = 1, \dots, n$  and  $h = 1, \dots, k$ , corresponding to aggregate  $i$ . The first row of  $a_{\text{coeff}}$  gives the radii  $a_{0,0}^{(i)}$  of the spheres associated with the set of aggregates.

Statistical analysis in [18] suggests to scale the aggregates by the radii of their first order spherical approximations and assume that (1) the scaled aggregates belong to the same population and (2) the properties of the scaled aggregates and the scale of these aggregates are independent random elements. Let  $Z_1 > 0$  be the random variable giving aggregate scale, so that the coefficients  $a_{0,0}^{(i)}$ ,  $i = 1, \dots, n_s$ , in the first row of  $a_{\text{coeff}}$  are interpreted as independent samples of  $Z_1$ .

Figure 6.10 shows estimates of the mean, standard deviation, skewness, and kurtosis for the scaled data  $h_n(\theta, \varphi)^{(i)} = g_n(\theta, \varphi)^{(i)} / a_{0,0}^{(i)}$ ,  $i = 1, \dots, n_s$ , with  $n = 30$  in the top left, top right, bottom left, and bottom right panels as functions of the arguments  $(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]$ . Since the estimates vary with  $(\theta, \varphi)$ , skewness is positive, and kurtosis differs from 3, a non-Gaussian inhomogeneous random field is needed to capture aggregate geometry.

### 6.3.2.3 Model Calibration

Let  $(\Omega, \mathcal{F}, P)$  be a probability space, where the sample space  $\Omega$  consists of all possible outcomes, that is, all possible virtual aggregates, the  $\sigma$ -field  $\mathcal{F}$  collects all relevant events, and  $P$  is a probability measure defined on  $\mathcal{F}$ . Let



**Fig. 6.10** Estimates of mean, standard deviation, skewness, and kurtosis for a population of  $n_s = 128$  scaled aggregates

$$H : [0, \pi] \times [0, 2\pi] \times \Omega \rightarrow [0, \infty) \quad (6.32)$$

be a random field defined on this probability space, whose samples constitute virtual aggregates. Since  $H(\theta, \varphi, \cdot) : \Omega \rightarrow [0, \infty)$  is a random variable for a fixed  $(\theta, \varphi)$  giving centroid to boundary point  $(\theta, \varphi)$  distances, it cannot take negative values. The functions  $h^{(i)}(\theta, \varphi)$ ,  $i = 1, \dots, n_s$ , are viewed as independent samples of  $H(\theta, \varphi)$ , so that  $H$  models the scaled aggregates and

$$G(\theta, \varphi) = Z_1 H(\theta, \varphi), \quad (\theta, \varphi) \in [0, \pi] \times [0, 2\pi], \quad (6.33)$$

models actual size aggregates. Recall that  $Z_1$  is assumed to be independent of  $H$ .

We approximate  $H$  by a sequence of translation random fields  $\{H_{T,n}\}$ ,  $n = 0, 1, \dots$ , defined by

$$H_{T,n}(\theta, \varphi) = F^{-1}\left(\Phi(N_n(\theta, \varphi)); \theta, \varphi\right) \quad n = 0, 1, \dots, \quad (6.34)$$

where  $F(\cdot; \theta, \varphi)$  denotes the marginal distribution of  $H$  at  $(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]$  and  $\{N_n(\theta, \varphi)\}$ ,  $n = 0, 1, \dots$ , is the sequence of Gaussian fields with mean 0, variance 1, and covariance function

$$\rho_n((\theta, \varphi), (\theta', \varphi')) = E [N_n(\theta, \varphi) N_n(\theta', \varphi')]. \quad (6.35)$$

The random fields  $H_{T,n}$  are completely defined by the mapping  $q = F^{-1} \circ \Phi$  and the covariance function  $\rho_n$  of  $N_n$ .

The construction of the Gaussian field  $N_n$  in (6.34) can be based on spherical harmonics. Let  $\{X_n(\theta, \varphi)\}$ ,  $n = 0, 1, \dots$ ,  $(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]$ , be a sequence of Gaussian random fields defined by

$$X_n(\theta, \varphi) = \sum_{k=0}^n \left[ \tilde{A}_{k,0}^{(g)} p_k(\cos(\theta)) + \sum_{h=1}^k \left( \tilde{A}_{k,h}^{(g)} \cos(h\varphi) + \tilde{B}_{k,h}^{(g)} \sin(h\varphi) \right) p_{k,h}(\cos(\theta)) \right], \quad (6.36)$$

where  $(\tilde{A}_{k,0}^{(g)}, \tilde{A}_{k,h}^{(g)}, \tilde{B}_{k,h}^{(g)})$  are Gaussian random variables and  $\tilde{A}_{0,0}^{(g)} = 1$ . The normalized version,

$$N_n(\theta, \varphi) = \frac{X_n(\theta, \varphi) - E[X_n(\theta, \varphi)]}{\sqrt{E[(X_n(\theta, \varphi) - E[X_n(\theta, \varphi)])^2]}}, \quad (6.37)$$

of  $X_n$  is a Gaussian random field with mean 0, variance 1, and covariance function  $\rho_n((\theta, \varphi), (\theta', \varphi'))$ . Numerical experiments suggest that the scaled covariance function of  $H(\theta, \varphi)$  does not differ significantly from  $\rho_n$  ([12], Sect. 3.1.1). For simplicity, we approximate  $\rho_n$  by this scaled covariance function. In summary, the random fields  $H_{T,n}$  in (6.34) have the required marginal distribution for all  $n \geq 0$  and their covariance function differs slightly from that of  $H$ . The random field  $G(\theta, \varphi)$  in (6.33) is approximated by the translation model

$$G_{T,n}(\theta, \varphi) = Z_1 H_{T,n}(\theta, \varphi), \quad (\theta, \varphi) \in [0, \pi] \times [0, 2\pi], \quad (6.38)$$

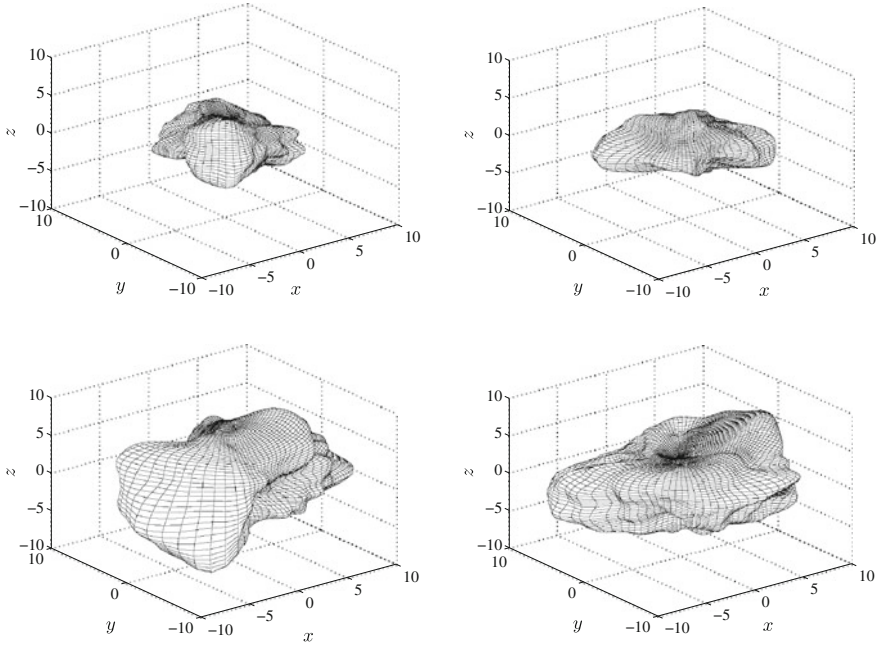
with  $H_{T,n}$  in (6.34). Note that the model in (6.38) is physically consistent in the sense that  $G_{T,n}(\theta, \varphi) > 0$  a.s. The distribution of  $Z_1$  can be estimated from its samples in the first row of  $a_{\text{coeff}}$ . The marginal distribution and the covariance functions of  $H_{T,n}$  can be estimated from rows 2 to 961 of  $a_{\text{coeff}}$  scaled by the first row of this matrix.

The model  $G_{T,n}(\theta, \varphi)$  calibrated to the data set in matrix  $a_{\text{coeff}}$  can be used to generate virtual aggregates that are statistically consistent with the available population. The generation algorithm involves two steps. First, samples of  $N_n(\theta, \varphi)$  and  $Z_1$  need to be generated. Second, samples of  $G_{T,n}(\theta, \varphi)$  need to be calculated from (6.38) and samples of  $N_n(\theta, \varphi)$  and  $Z_1$ . Figure 6.11 shows four virtual aggregates, that is, four samples of  $G_{T,n}(\theta, \varphi)$  in (6.38) with  $n = 30$ .

### 6.3.3 Probabilistic Models for Microstructures

Consider the stochastic elliptic boundary value problem

$$\begin{aligned} -\nabla \cdot (A(x) \nabla U(x)) &= b(x), \quad x \in D \\ U(x) &= 0, \quad x \in \partial D, \end{aligned} \quad (6.39)$$



**Fig. 6.11** Virtual aggregates given by spherical harmonic representations with  $n = 30$

where  $D$  is a bounded subset of  $\mathbb{R}^d$ ,  $A(x)$ ,  $x \in D$ , is a real-valued random field on a probability space  $(\Omega, \mathcal{F}, P)$ , and  $b(x)$ ,  $x \in D$ , denotes a real-valued continuous function. We discuss stochastic equations of the type in (6.39) in Sect. 9.4. Most studies of (6.39) focus on mathematical conditions that  $A(x)$  must satisfy such that this equation admits a unique weak solution [1–3, 7, 9, 10, 35]. Physical requirements on  $A(x)$  relate to, for example, features of microstructures that define the probability law of  $A(x)$  and can be inferred from data and/or any other available information ([19, 20] and [39], Chaps. 8 and 13), have not been examined systematically.

A key assumption in [1–3, 7, 9, 10, 35], referred to as the finite dimensional noise assumption, is that  $A(x)$  can be represented approximately by linear models of the type

$$A^{(n)}(x, \omega) = \sum_{k=1}^n C_k(\omega) \theta_k(x), \quad x \in D, \quad \omega \in \Omega, \quad (6.40)$$

where  $n \geq 1$  is an integer,  $\{C_k(\omega)\}$  are random variables on  $(\Omega, \mathcal{F}, P)$  that are usually assumed to be independent, and  $\{\theta_k(x)\}$  denote specified deterministic functions. In contrast to  $A(x)$  that usually consists of an uncountable number of random variables,  $A^{(n)}(x)$  depends on a finite number of random variables. Truncated Karhunen–Loève expansions and discrete spectral representations (Sects. 3.6.4, 3.6.5) are examples of linear models of the type in (6.40).



Let  $U^{(n)}(x)$  denote the solution of (6.39) with  $A^{(n)}(x)$  in place of  $A(x)$ , that is,

$$\begin{aligned} -\nabla \cdot (A^{(n)}(x, y) \nabla U^{(n)}(x, y)) &= b(x) \quad (x, y) \in D \times \Gamma \\ U^{(n)}(x, y) &= 0 \quad (x, y) \in \partial D \times \Gamma, \end{aligned} \quad (6.41)$$

where  $\Gamma = (C_1, \dots, C_n)(\Omega) \subset \mathbb{R}^n$  denotes the image of the random variables in the expression of  $A^{(n)}(x)$ . Note that under the finite dimensional noise assumption the stochastic partial differential equation in (6.39) becomes the deterministic parametric elliptic partial differential equation in (6.41) defined on  $D \times \Gamma$  [3]. The measure on  $D \times \Gamma$  is the product of the Lebesgue measure on  $D$  and the probability measure  $f(y) dy$  on  $\Gamma$ , where  $f(y)$  denotes the joint probability density function of  $(C_1, \dots, C_n)$ . To assure the existence and uniqueness of the solution of (6.41), the finite dimensional noise assumption is supplemented by additional constraints on the samples of  $A^{(n)}(x)$  [1–3, 7, 9, 10, 15, 35].

We examine potential limitations of the linear model in (6.40) with independent random coefficients  $\{C_k\}$ , and develop alternative models for  $A(x)$ . It is shown that linear models with independent random coefficients of the type in (6.40) are approximately Gaussian, so that they may be unsatisfactory in many applications, and that efficient algorithms can be developed for constructing linear models that are consistent with both mathematical and physical constraints.

### 6.3.3.1 Linear Models with Independent Coefficients

Suppose  $A(x)$  in (6.39) is a real-valued random field with correlation function that is continuous and square integrable in  $D \times D$ . Then  $A(x)$  admits the Karhunen–Loève expansion

$$A(x, \omega) = E[A(x)] + \sum_{k=1}^{\infty} \lambda_k^{1/2} \theta_k(x) Y_k(\omega), \quad x \in D, \quad (6.42)$$

where the equality is understood in m.s.,  $\{Y_k\}$  are uncorrelated random variables with mean 0 and variance 1, and  $\{\lambda_k, \theta_k(x)\}$  denote the eigenvalues and eigenfunctions of the correlation function of  $A(x)$  ([21], Sect. 6.2, and Sect. 3.6.5 in this book). Truncations,

$$A^{(n)}(x, \omega) = E[A(x, \omega)] + \sum_{k=1}^n \lambda_k^{1/2} \theta_k(x) Y_k(\omega), \quad x \in D, \quad (6.43)$$

of the representation of  $A(x)$  in (6.42) produce linear models for  $A(x)$  of the type in (6.40). Most developments on stochastic Galerkin and collocation methods assume that the random variables  $\{Y_k\}$  in (6.42) and (6.43) are independent and take values in bounded intervals. The assumption of independence is valid for Gaussian fields but is invalid for non-Gaussian random fields. If  $A(x)$  is a Gaussian field, the random variables  $\{Y_k\}$  are Gaussian so that they cannot be bounded.

Linear models similar to those in (6.43) can be obtained by the spectral representation theorem for weakly stationary random fields ([13] and Sect. 3.6.4 in this book). These models are trigonometric polynomials with random coefficients. Since our discussion is limited to homogeneous random fields  $A(x)$ , we only consider linear models derived from the spectral representation theorem. For simplicity, it is assumed that  $D$  in (6.40) is an interval of the real line.

Let  $A(x)$  be a real-valued weakly stationary random field with mean 0, variance 1, and one-sided spectral density  $g(\nu)$  with frequency band  $(0, \bar{\nu})$ ,  $0 < \bar{\nu} < \infty$ . Define the sequence of random fields

$$A^{(n)}(x) = \sum_{k=1}^n \sigma_k^{(n)} \left( A_{c,k}^{(n)} \cos(v_k^{(n)} x) + A_{s,k}^{(n)} \sin(v_k^{(n)} x) \right), \quad x \in D, \quad n = 1, 2, \dots, \quad (6.44)$$

where  $\{A_{c,k}^{(n)}, A_{s,k}^{(n)}\}$  are uncorrelated random variables with mean 0 and variance 1,  $\Delta\nu = \bar{\nu}/n$ ,  $v_k^{(n)} = (k - 1/2)\Delta\nu$ , and  $(\sigma_k^{(n)})^2 = \int_{(k-1)\Delta\nu}^{k\Delta\nu} g(\nu) d\nu$ ,  $k = 1, \dots, n$ . The mean and covariance functions of  $A^{(n)}(x)$  are  $E[A^{(n)}(x)] = 0$  and  $E[A^{(n)}(x)A^{(n)}(x')] = \sum_{k=1}^n (\sigma_k^{(n)})^2 \cos(v_k^{(n)}(x - x'))$ , so that  $A^{(n)}(x)$  has unit variance. The first two moments of  $A^{(n)}(x)$  converge to those of  $A(x)$  as  $n \rightarrow \infty$  for  $\bar{\nu} < \infty$ , and this convergence extends to the case in which  $A(x)$  has an unbounded frequency band (Sect. 3.8.1). If  $A(x)$  is a Gaussian field, then  $\{A_{c,k}^{(n)}, A_{s,k}^{(n)}\}$  are independent  $N(0, 1)$  variables. Otherwise,  $\{A_{c,k}^{(n)}, A_{s,k}^{(n)}\}$  are uncorrelated but dependent random variables.

Suppose, following current developments, that  $\{A_{c,k}^{(n)}, A_{s,k}^{(n)}\}$  are independent random variables with specified distributions. The characteristic function of  $A^{(n)}(x)$  is

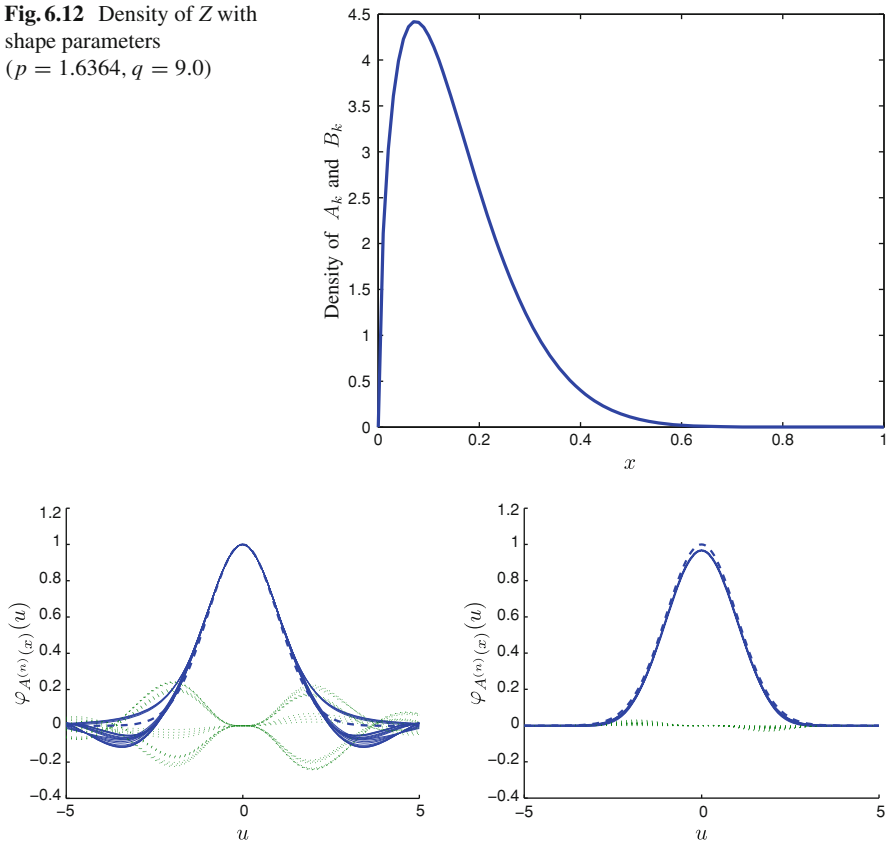
$$\varphi_{A^{(n)}(x)}(u) = E \left[ e^{iuA^{(n)}(x)} \right] = \prod_{k=1}^n \left( \varphi_{A_{c,k}^{(n)}}(u\sigma_k^{(n)} \cos(v_k^{(n)} x)) \varphi_{A_{s,k}^{(n)}}(u\sigma_k^{(n)} \sin(v_k^{(n)} x)) \right), \quad (6.45)$$

where  $\varphi_{A_{c,k}^{(n)}}$  and  $\varphi_{A_{s,k}^{(n)}}$  denote the characteristic functions of  $A_{c,k}^{(n)}$  and  $A_{s,k}^{(n)}$ .

**Example 6.3** Suppose  $\{A_{c,k}^{(n)}, A_{s,k}^{(n)}\}$  are independent random variables following a Beta distribution with shape parameters  $(p > 0, q > 0)$  and range  $[a, b]$ , so that  $A_{c,k}^{(n)} \stackrel{d}{=} A_{s,k}^{(n)} \stackrel{d}{=} Y$ ,  $Y = a + (b - a)Z$ , and  $Z$  is a standard Beta variable with parameters  $(p, q)$ . The density of  $Z$  is  $f(z) = z^{p-1}(1 - z)^{q-1}/B(p, q)$ ,  $z \in [0, 1]$ , with mean  $E[Z] = p/(p + q)$  and variance  $\text{Var}[Z] = pq/[(p + q)^2(p + q + 1)]$  ([25], Chap. 24). If the parameters  $(a, b, p, q)$  satisfy the conditions  $a = -(p/q)b$ ,  $p = q(q + 1)/(b^2 - q)$ , and  $b^2 > q$ , the random variable  $Y$  has mean 0 and variance 1, for example,  $E[Y] = 0$  and  $E[Y^2] = 1$  for  $(a = -1.4545, b = 8, p = 1.6364, q = 9)$ . The density of the standard Beta variable  $Z$  for these parameters is highly skewed to the left, as illustrated in Fig. 6.12.

The characteristic function of  $A^{(n)}(x)$  is given by (6.45) with  $\varphi_{A_{c,k}^{(n)}}$  and  $\varphi_{A_{s,k}^{(n)}}$  equal to the characteristic function of  $Y$ . Figure 6.13 shows with heavy dotted lines

**Fig. 6.12** Density of  $Z$  with shape parameters ( $p = 1.6364, q = 9.0$ )



**Fig. 6.13** Real and imaginary parts of the characteristic functions  $\varphi_{A^{(n)}(x)}(u)$  are shown with *thin solid* and *thin dotted* lines, respectively, for  $n = 1$  (left panel) and  $n = 10$  (right panel). *Heavy dotted lines* are the characteristic function of  $N(0, 1)$

the characteristic function of the standard normal variable. The thin solid and dotted lines in the figure are real and imaginary parts of  $\varphi_{A^{(n)}(x)}(u)$  at 20 equally spaced arguments  $x \in [0, \tau = 10]$  for  $n = 1$  (left panel) and  $n = 10$  (right panel). As  $n$  increases, the marginal distribution of  $A^{(n)}(x)$  approaches rapidly the distribution of  $N(0, 1)$  and becomes space invariant.  $\diamond$

The numerical experiments in Example 6.3 are consistent with the following two theorems showing that  $A^{(n)}(x)$  becomes a stationary Gaussian random function as  $n \rightarrow \infty$  under some mild conditions on  $A_{c,k}^{(n)}$  and  $A_{s,k}^{(n)}$ . Consider the alternative form,

$$A^{(n)}(x) = \sum_{k=1}^n (U_k^{(n)}(x) + V_k^{(n)}(x)), \quad (6.46)$$

of the linear model  $A^{(n)}(x)$  in (6.44), where  $U_k^{(n)}(x) = \sigma_k^{(n)} A_{c,k}^{(n)} \cos(v_k^{(n)} x)$  and  $V_k^{(n)}(x) = \sigma_k^{(n)} A_{s,k}^{(n)} \sin(v_k^{(n)} x)$ . Let  $F_{U_k^{(n)}(x)}$  and  $F_{V_k^{(n)}(x)}$  denote the distributions of random variables  $U_k^{(n)}(x)$  and  $V_k^{(n)}(x)$ . These distributions have means 0 and variances  $(\sigma_k^{(n)} \cos(v_k^{(n)} x))^2 \sim O(n^{-1})$  and  $(\sigma_k^{(n)} \sin(v_k^{(n)} x))^2 \sim O(n^{-1})$ , respectively. Note that the random variables  $U_k^{(n)}(x)$  and  $V_k^{(n)}(x)$  are uncorrelated,  $E[A^{(n)}(x)] = 0$ , and  $E[A^{(n)}(x)^2] = 1$ .

**Theorem 6.1** *If  $\{A_{c,k}^{(n)}, A_{s,k}^{(n)}\}$  in (6.44) are independent random variables and the distributions  $F_{U_k^{(n)}(x)}$  and  $F_{V_k^{(n)}(x)}$  are such that  $\int_{|z|>\xi} z^2 F_{U_k^{(n)}(x)}(dz) \sim O(n^{-(1+\alpha)})$  and  $\int_{|z|>\xi} z^2 F_{V_k^{(n)}(x)}(dz) \sim O(n^{-(1+\alpha)})$  hold as  $n \rightarrow \infty$  for  $\alpha > 0$  and arbitrary  $\xi > 0$ , then  $A^{(n)}(x) \Rightarrow N(0, 1)$  at each  $x \in D$  as  $n \rightarrow \infty$ .*

*Proof* Set  $S_n = \sum_{k=1}^n Z_k$  and  $s_n^2 = \sum_{k=1}^n \rho_k^2$ , where  $\{Z_k, k \geq 1\}$  are independent random variables with means  $E[Z_k] = 0$ , variances  $\rho_k^2 = E[Z_k^2]$ , and distributions  $F_k$ . If the Lindeberg–Feller condition is satisfied, that is,

$$\frac{1}{s_n^2} \sum_{k=1}^n E[Z_k^2 1(|Z_k/s_n| > \xi)] = \frac{1}{s_n^2} \sum_{k=1}^n \int_{|z|>\xi s_n} z^2 F_k(dz) \rightarrow 0, \quad \text{as } n \rightarrow \infty, \quad (6.47)$$

for all  $\xi > 0$ , then  $S_n/s_n \Rightarrow N(0, 1)$  as  $n \rightarrow \infty$ , that is, the sequence of distributions of  $S_n/s_n$  converges to the distribution of the standard Gaussian variable  $N(0, 1)$  as  $n$  increases indefinitely ([33], Theorem 9.8.1).

We show that the Lindeberg–Feller condition holds for the sequence of independent random variables  $\{U_k^{(n)}(x), V_k^{(n)}(x)\}$  in the definition of  $A^{(n)}(x)$ . For our case, the parameter  $s_n^2$  and the integrals  $\int_{|z|>\xi s_n} z^2 F_k(dz)$  in (6.47) are  $E[A^{(n)}(x)^2] = 1$  and  $\int_{|z|>\xi} z^2 F_{U_k^{(n)}(x)}(dz)$ ;  $\int_{|z|>\xi} z^2 F_{V_k^{(n)}(x)}(dz)$ , respectively. Accordingly, the summation  $(1/s_n^2) \sum_{k=1}^n \int_{|z|>\xi s_n} z^2 F_k(dz)$  is of order  $\sum_{k=1}^n (O(n^{-(1+\alpha)}) + O(n^{-(1+\alpha)})) = O(n^{-\alpha})$ , so that  $A^{(n)}(x) \Rightarrow N(0, 1)$  as  $n \rightarrow \infty$  at each  $x \in D$ . The assumption that  $A(x)$  has a bounded frequency band  $[0, \bar{\nu}]$ ,  $\bar{\nu} < \infty$ , can be relaxed since our arguments hold for an arbitrary  $\bar{\nu}$  so that they also apply in the limit as  $\bar{\nu} \rightarrow \infty$ .  $\blacktriangle$

Note that if the random variables  $(A_{c,k}^{(n)}, A_{s,k}^{(n)})$  in (6.44) are bounded a.s., that is, there exists  $a > 0$  finite such that  $P(|A_{c,k}^{(n)}| > a) = P(|A_{s,k}^{(n)}| > a) = 0$  implying  $|U_k^{(n)}(x)| \leq \sigma_k^{(n)} a$  and  $|V_k^{(n)}(x)| \leq \sigma_k^{(n)} a$  a.s. Since  $\sigma_k^{(n)} \sim O(n^{-1/2})$ , there exists an integer  $n_\xi \geq 1$  for  $\xi > 0$  such that  $\sigma_k^{(n)} a < \xi$ ,  $n \geq n_\xi$ . Hence,  $A^{(n)}(x) \Rightarrow N(0, 1)$  as  $n \rightarrow \infty$  by the Lindeberg–Feller condition, since the integrals  $\int_{|z|>\xi} z^2 F_{U_k^{(n)}(x)}(dz)$  and  $\int_{|z|>\xi} z^2 F_{V_k^{(n)}(x)}(dz)$  are equal to 0 for  $n \geq n_\xi$ .

**Theorem 6.2** *Under the conditions of Theorem 6.1,  $A^{(n)}(x)$  converges to a stationary Gaussian random function as  $n \rightarrow \infty$  with the second moment properties of  $A(x)$ .*

*Proof* We need to show that for any integer  $m \geq 1$ , arguments  $x_1, \dots, x_m \in D$ , and coefficients  $\beta_1, \dots, \beta_m \in \mathbb{R}$ , the random variable

$$S_{m,n} = \sum_{i=1}^m \beta_i A^{(n)}(x_i) = \sum_{k=1}^n \sum_{i=1}^m \beta_i (U_k^{(n)}(x_i) + V_k^{(n)}(x_i))$$

is Gaussian ([23], Sect. 5.5). The mean and variance of  $S_{m,n}$  are  $E[S_{m,n}] = \sum_{i=1}^m \beta_i E[A^{(n)}(x_i)] = 0$  and

$$\begin{aligned} E[S_{m,n}^2] &= \sum_{i,j=1}^m \beta_i \beta_j E[A^{(n)}(x_i) A^{(n)}(x_j)] = \sum_{i,j=1}^m \beta_i \beta_j \sum_{k=1}^n (\sigma_k^{(n)})^2 \cos(v_k^{(n)}(x_i - x_j)) \\ &\leq \sum_{i,j=1}^m |\beta_i \beta_j| \sum_{k=1}^n (\sigma_k^{(n)})^2 |\cos(v_k^{(n)}(x_i - x_j))| \leq \left( \sum_{i=1}^m |\beta_i| \right)^2 < \infty \end{aligned}$$

since  $\sum_{k=1}^n (\sigma_k^{(n)})^2 = 1$  by assumption. Note that  $\{\beta_i U_k^{(n)}(x_i)\}$  and  $\{\beta_i V_k^{(n)}(x_i)\}$  play the role of the random variables  $\{Z_k\}$  in (6.47). The Lindeberg–Feller condition requires that

$$\begin{aligned} \frac{1}{s_{m,n}^2} \sum_{k=1}^n \sum_{i=1}^m \beta_i^2 &\left[ \int_{|U_k^{(n)}(x_i)| > \xi s_{m,n}/|\beta_i|} z^2 F_{U_k^{(n)}(x_i)}(dz) \right. \\ &\left. + \int_{|V_k^{(n)}(x_i)| > \xi s_{m,n}/|\beta_i|} z^2 F_{V_k^{(n)}(x_i)}(dz) \right] \end{aligned}$$

converges to 0 as  $n \rightarrow \infty$  for all  $\xi > 0$ . Under the conditions of the previous theorem, we have this convergence so that  $A^{(n)}(x)$  becomes a Gaussian function for large values of  $n$ . Since the second moment properties of  $A^{(n)}(x)$  in (6.44) converge to those of  $A(x)$ , the limit of  $A^{(n)}(x)$  as  $n \rightarrow \infty$  is a Gaussian field that is equal to  $A(x)$  in the second moment sense.  $\blacktriangle$

### 6.3.3.2 Linear Models with Dependent Coefficients

Let  $A(x)$ ,  $x \in D$ , be a real-valued random field with finite variance defined on a bounded interval  $D$  of the real line, and let  $A^{(n)}(x) = \sum_{k=1}^n C_k \theta_k(x)$  be a linear model of the type in (6.40). The coefficients  $\{C_k\}$  are uncorrelated but dependent random variables, unless  $A(x)$  is a stationary Gaussian field. As previously mentioned, most methods for solving stochastic elliptic partial differential equations assume that  $\{C_k\}$  are independent random variables. There may be two reasons for this assumption. First, the construction of the joint distribution of  $(C_1, \dots, C_n)$  such that the probability law of  $A^{(n)}(x)$  matches that of a target random field  $A(x)$  is difficult. Second, the calculation of expectations of the type  $E[h(C_1, \dots, C_n)]$  involved in numerical solutions of (6.39) and (6.41) is simpler if  $\{C_k\}$  are independent rather than dependent random variables, where  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  denotes a measurable function.

Let  $A(x, \omega_s), \omega_s \in \Omega, s = 1, \dots, n_s$ , be  $n_s$  independent samples of  $A(x), x \in D$ . Our objective is to construct samples  $(C_1(\omega_s), \dots, C_n(\omega_s))$  of  $(C_1, \dots, C_n)$  such that the corresponding samples  $A^{(n)}(x, \omega_s) = \sum_{k=1}^n C_k(\omega_s) \theta_k(x), s = 1, \dots, n_s$ , of  $A^{(n)}(x)$  are close to the samples  $A(x, \omega_s), s = 1, \dots, n_s$ , of  $A(x)$ . For example, we may select  $(C_1(\omega_s), \dots, C_n(\omega_s))$  to minimize a distance  $d(A(x, \omega_s), A^{(n)}(x, \omega_s)), s = 1, \dots, n_s$ , between samples of  $A(x)$  and  $A^{(n)}(x)$ . If  $A(x)$  has continuous samples and the functions  $\{\theta_k(x)\}$  are continuous, the discrepancy between  $A(x, \omega_s)$  and  $A^{(n)}(x, \omega_s)$  can be measured by

$$d(A(x, \omega_s), A^{(n)}(x, \omega_s)) = \sup_{x \in D} \left| A(x, \omega_s) - \sum_{k=1}^n C_k(\omega_s) \theta_k(x) \right|, \quad (6.48)$$

that is, the sup metric in the space  $C(D)$  of real-valued continuous functions defined on  $D$ . Optimization algorithms can be used to find  $\{C_k(\omega_s)\}$  minimizing the objective function in (6.48). Alternative objective functions can be used to construct samples of  $(C_1, \dots, C_n)$ , for example, objective functions depending on differences between samples of  $A(x)$  and  $A^{(n)}(x)$  and of  $dA(x)/dx$  and  $dA^{(n)}(x)/dx$  provided that almost all samples of  $A(x)$  and  $dA(x)/dx$  and the derivatives of the basis functions  $\{\theta_k(x)\}$  are continuous functions. The optimization algorithm has a unique solution. For example, if the basis functions  $\{\theta_k(x)\}$  are algebraic or trigonometric polynomials, then for every  $\varepsilon > 0$  there exists an integer  $n(\varepsilon, \omega_s) \geq 1$  such that  $d(A(x, \omega_s), A^{(n(\varepsilon, \omega_s))}(x, \omega_s)) < \varepsilon$  ([4], Theorem 6.1.1). Moreover, the coefficients  $(C_1(\omega_s), \dots, C_n(\omega_s))$  of the polynomial with this property are unique ([4], Corollary 7.4.2). Similar results can be found elsewhere ([30], Theorem 13.1 and Sect. 14.3).

Let  $(C_1(\omega_s), \dots, C_n(\omega_s)), s = 1, \dots, n_s$ , be  $n_s$  samples of  $(C_1, \dots, C_n)$  corresponding to  $n_s$  independent samples  $A(x, \omega_s), s = 1, \dots, n_s$ , of a target random field  $A(x)$  delivered by an optimization algorithm with objective function of the type in (6.48). The samples of  $(C_1, \dots, C_n)$  can be stored for later processing and/or used to estimate statistics of this random vector. We construct estimators for the characteristic and the distribution functions of  $(C_1, \dots, C_n)$  and of the linear model  $A^{(n)}(x)$ .

Let

$$\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n) = \frac{1}{n_s} \sum_{s=1}^{n_s} e^{i \sum_{k=1}^n u_k C_k^{(s)}}, \quad (u_1, \dots, u_n) \in \mathbb{R}^n, \quad (6.49)$$

be an estimator of the characteristic function

$$\varphi_{C_1, \dots, C_n}(u_1, \dots, u_n) = E[e^{i \sum_{k=1}^n u_k C_k}], \quad (u_1, \dots, u_n) \in \mathbb{R}^n, \quad (6.50)$$

of the random vector  $(C_1, \dots, C_n)$ , where  $(C_1^{(s)}, \dots, C_n^{(s)}), s = 1, \dots, n_s$ , are independent copies of this vector. Let also

$$\hat{\varphi}_{A^{(n)}(x_1), \dots, A^{(n)}(x_m)}(u_1, \dots, u_m) = \hat{\varphi}_{C_1, \dots, C_n} \left( \sum_{r=1}^m u_r \theta_1(x_r), \dots, \sum_{r=1}^m u_r \theta_n(x_r) \right), \quad (6.51)$$

be an estimator of the characteristic function

$$\begin{aligned}\varphi_{A^{(n)}(x_1), \dots, A^{(n)}(x_m)}(u_1, \dots, u_m) &= E\left[e^{i \sum_{r=1}^m u_r A^{(n)}(x_r)}\right] \\ &= \varphi_{C_1, \dots, C_n}\left(\sum_{r=1}^m u_r \theta_1(x_r), \dots, \sum_{r=1}^m u_r \theta_n(x_r)\right)\end{aligned}\quad (6.52)$$

of  $(A^{(n)}(x_1), \dots, A^{(n)}(x_m))$ , where  $m \geq 1$  is an integer and  $x_1, \dots, x_m \in D$  are distinct points. Similar estimators can be constructed for the distribution of  $(C_1, \dots, C_n)$ .

**Theorem 6.3** *The estimators  $\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n)$  and  $\hat{\varphi}_{A^{(n)}(x_1), \dots, A^{(n)}(x_m)}(u_1, \dots, u_m)$  in (6.49) and (6.51) are unbiased and weakly consistent. The discrepancy between the characteristic functions of  $A(x)$  and  $A^{(n)}(x)$  can be bounded by*

$$E[|\varphi_A(u) - \hat{\varphi}_{A^{(n)}(x)}(u)|] \leq |\varphi_A(u) - \varphi_{A^{(n)}(x)}(u)| + (\text{Var}[\hat{\varphi}_{A^{(n)}(x)}(u)])^{1/2}. \quad (6.53)$$

*Proof* We show that the mean of  $\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n)$  is  $\varphi_{C_1, \dots, C_n}(u_1, \dots, u_n)$  and that

$$P(|\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n) - \varphi_{C_1, \dots, C_n}(u_1, \dots, u_n)| > \varepsilon) \rightarrow 0$$

as  $n_s \rightarrow \infty$  for arbitrary  $\varepsilon > 0$  ([33], Sect. 6.2.1). Similar conditions must be satisfied by the estimator in (6.51). Elementary calculations give

$$\begin{aligned}E[\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n)] &= \varphi_{C_1, \dots, C_n}(u_1, \dots, u_n) \\ \text{Var}[\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n)] &= \frac{1}{n_s} [1 - |\varphi_{C_1, \dots, C_n}(u_1, \dots, u_n)|^2] \leq \frac{1}{n_s},\end{aligned}\quad (6.54)$$

so that the estimator  $\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n)$  is unbiased. The Chebyshev inequality,

$$\begin{aligned}P(|\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n) - \varphi_{C_1, \dots, C_n}(u_1, \dots, u_n)| > \varepsilon) \\ \leq \text{Var}[\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n)] / \varepsilon^2,\end{aligned}$$

implies that the estimator is weakly consistent. Since the estimator in (6.51) is the estimator  $\hat{\varphi}_{C_1, \dots, C_n}(u_1, \dots, u_n)$  for particular arguments, it has the same properties as this estimator.

The discrepancy between the characteristic function  $\varphi_{A(x)}(u) = \varphi_A(u)$  of  $A(x)$  and the estimator of the marginal characteristic function of  $A^{(n)}(x)$  in (6.52) for  $m = 1$  can be bounded by

$$|\varphi_A(u) - \hat{\varphi}_{A^{(n)}(x)}(u)| \leq |\varphi_A(u) - \varphi_{A^{(n)}(x)}(u)| + |\varphi_{A^{(n)}(x)}(u) - \hat{\varphi}_{A^{(n)}(x)}(u)|, \quad (6.55)$$

so that

$$E[|\varphi_A(u) - \hat{\varphi}_{A^{(n)}(x)}(u)|] \leq |\varphi_A(u) - \varphi_{A^{(n)}(x)}(u)| + E[|\varphi_{A^{(n)}(x)}(u) - \hat{\varphi}_{A^{(n)}(x)}(u)|]. \quad (6.56)$$

The bound in (6.53) results by applying the Cauchy-Schwarz inequality to the second term on the right side of (6.56). The first and the second terms of the bound

on  $E[|\varphi_A(u) - \hat{\varphi}_{A^{(n)}(x)}(u)|]$  relate to model accuracy and statistical uncertainty, respectively. Since  $\text{Var}[\hat{\varphi}_{A^{(n)}(x)}(u)] \leq 1/n_s$ , the second term can be made as small as desired by increasing the sample size.  $\blacktriangle$

Additional information on sample properties of  $A(x)$  is needed to bound the discrepancy  $|\varphi_A(u) - \varphi_{A^{(n)}(x)}(u)|$  corresponding to model accuracy (Theorem 6.4). Suppose

$$A^{(n)}(x) = \frac{A_0^{(n)}}{2} + \sum_{k=1}^n (A_k^{(n)} \cos(\nu_k x) + B_k^{(n)} \sin(\nu_k x)), \quad t \in [0, T], \quad (6.57)$$

where  $\{A_k^{(n)}, k = 0, 1, \dots, n, B_k^{(n)}, k = 1, \dots, n\}$  are random variables. The random variables  $\{A_k^{(n)}, B_k^{(n)}\}$  and the basis functions  $(1, \dots, \cos(\nu_k x), \dots, \sin(\nu_k x), \dots)$  in (6.57) are similar to  $\{C_k\}$  and  $\{\theta_k(x)\}$  in (6.40). The samples of  $A^{(n)}(x)$  are members of the linear space

$$\mathcal{P}_n[0, T] = \left\{ q(t) = \frac{\alpha_0}{2} + \sum_{k=1}^n (\alpha_k \cos(\nu_k x) + \beta_k \sin(\nu_k x)) \right\} \quad (6.58)$$

of trigonometric polynomials of order  $n$  with period  $T > 0$ , where  $\nu_1 = 2\pi/T$ ,  $\nu_k = k\nu_1$ , and  $\alpha_k, \beta_k \in \mathbb{R}$ ,  $k = 1, 2, \dots, n$ . There are at least two reasons for focussing on linear models of the type in (6.57). First, trigonometric polynomials have been and are used extensively in applications to approximate deterministic and random functions. Second, the model in (6.57) resembles the discrete spectral representation for weakly stationary random fields and can be used to represent both stationary and non-stationary non-Gaussian random fields [16].

The following algorithm can be applied to find samples  $\{A_k^{(n)}(\omega), B_k^{(n)}(\omega)\}$  of  $\{A_k^{(n)}, B_k^{(n)}\}$  corresponding to samples  $A(x, \omega)$  of  $A(x)$ . Let  $q^* \in \mathcal{P}_n[0, T]$  be the optimal trigonometric polynomial of order  $n$  corresponding to a sample  $A(x, \omega)$  of  $A(x)$ , that is,  $q^*(x)$  has the property  $\|A(\cdot, \omega) - q^*(\cdot)\|_\infty = \min_{q \in \mathcal{P}_n[0, T]} \|A(\cdot, \omega) - q(\cdot)\|_\infty$ . The samples  $\{A_k^{(n)}(\omega), B_k^{(n)}(\omega)\}$  of  $\{A_k^{(n)}, B_k^{(n)}\}$  are set equal to the coefficients  $\{\alpha_k^*, \beta_k^*\}$  of  $q^*(x)$ .

**Theorem 6.4** *If almost all samples of  $A(x)$  are continuous and periodic with period  $T > 0$ , almost all samples of  $A'(x) = dA(x)/dx$  are continuous, and the random variable  $\|A'(\cdot, \omega)\|_\infty$  is integrable, then*

$$\begin{aligned} \|A(\cdot, \omega) - A^{(n)}(\cdot, \omega)\|_\infty &\leq \frac{\pi}{2(n+1)} \|A'(\cdot, \omega)\|_\infty, \\ |\varphi_{A(x)}(u) - \varphi_{A^{(n)}(x)}(u)| &\leq \frac{\pi|u|}{2(n+1)} E[\|A'(\cdot, \omega)\|_\infty], \end{aligned} \quad (6.59)$$

and

$$E[|\varphi_A(u) - \hat{\varphi}_{A^{(n)}(x)}(u)|] \leq \frac{\pi|u|}{2(n+1)} E[\|A'(\cdot, \omega)\|_\infty] + (\text{Var}[\hat{\varphi}_{A^{(n)}(x)}(u)])^{1/2}. \quad (6.60)$$



*Proof* We use a theorem for periodic functions with continuous derivatives stating that if  $h : \mathbb{R} \rightarrow \mathbb{R}$  is a continuous periodic function with period  $T$  and has continuous first order derivative, then  $\min_{q \in \mathcal{P}_n[0, T]} \|h - q\|_\infty \leq \pi \|h'\|_\infty / [2(n + 1)]$  ([30], Theorem 15.1). Since the samples  $A(x, \omega)$  of  $A(x)$  may not have the property  $A(0, \omega) = A(T, \omega)$ , the theorem cannot be applied directly. If  $A(0, \omega) \neq A(T, \omega)$ , there is no sequence of polynomials in  $\mathcal{P}_n[0, T]$  that converges to  $A(x, \omega)$  at  $x = 0$  and  $x = T$  ([38], Sect. 1.10). However,  $A(x, \omega)$  can be altered such that its modified version has equal values at the ends of  $[0, T]$ . Let

$$\begin{aligned} \tilde{A}(\tilde{x}, \omega) &= A(\tilde{x}, \omega) 1(0 \leq \tilde{x} \leq T - \varepsilon) \\ &\quad + [A(T, \omega) + (A(0, \omega) - A(T, \omega))(\tilde{x} - T + \varepsilon)/\varepsilon] 1(T - \varepsilon < \tilde{x} \leq T), \end{aligned} \quad (6.61)$$

where  $\varepsilon \in (0, T)$  is arbitrary and  $\tilde{x} = (1 - \varepsilon/T)x$ ,  $x \in [0, T]$ . We have  $\tilde{A}(0, \omega) = A(0, \omega)$ ,  $\tilde{A}(T - \varepsilon, \omega) = A(T, \omega)$ ,  $\tilde{A}(\tilde{x}, \omega)$  for  $\tilde{x} \in [0, T - \varepsilon]$  coincides with  $A(x, \omega)$  for  $x \in [0, T]$ ,  $\tilde{A}(T, \omega) = A(0, \omega)$ , and  $\tilde{A}(\tilde{x}, \omega)$  in  $[T - \varepsilon, T]$  is a line connecting the points  $\tilde{A}(T - \varepsilon, \omega) = A(T, \omega)$  and  $\tilde{A}(T, \omega) = A(0, \omega)$ . Higher order splines can be used to define  $\tilde{A}(\tilde{x})$  in  $[T - \varepsilon, 2T]$  such that, for example, its samples be continuously differentiable provided the samples of  $A(x)$  have this property. Similar arguments can be used for real-valued random fields defined on rectangles  $\times_{i=1}^d [0, T_i] \in \mathbb{R}^d$  by using properties of multivariate Fourier series ([38], Chap. 7). Alternatively, linear models  $A^{(n)}(x)$  can be constructed for  $A(x)$  on sets including  $D$  and the restriction of these models to  $D$  can be used for calculations.

Assume without loss of generality that  $A$  has the property  $A(0) = A(T)$  almost surely. The first inequality in (6.59) follows from [30] (Theorem 15.1). The second inequality in this equation follows from

$$\begin{aligned} |\varphi_{A(x)}(u) - \varphi_{A^{(n)}(x)}(u)| &= |E[e^{iuA(x)}(1 - e^{iu(A^{(n)}(x) - A(x))})]| \\ &\leq E[|1 - e^{iu(A^{(n)}(x) - A(x))}|] \\ &\leq |u| E[|A^{(n)}(x) - A(x)|] \leq \frac{\pi |u|}{2(n + 1)} E[\|A'(\cdot, \omega)\|_\infty], \end{aligned}$$

where we use the first inequality in (6.59) and  $|1 - \exp(ia)| \leq |a|$ ,  $a \in \mathbb{R}$ . The latter inequality and the bound in (6.53) give (6.60). The resulting bound shows that for a fixed  $u \in \mathbb{R}$  the discrepancy between a target characteristic function  $\varphi_A(u)$  and its estimate  $\hat{\varphi}_{A^{(n)}(x)}(u)$  corresponding to a linear model  $A^{(n)}(x)$  of order  $n$  and based on  $n_s$  independent samples of  $A(x)$  can be made as small as desired by increasing the model order and the sample size. Similar arguments can be used to construct bounds on the discrepancy between estimators for finite dimensional characteristic functions of  $A(x)$  and  $A^{(n)}(x)$ .  $\blacktriangle$

*Example 6.4* Let  $G(x)$ ,  $x \in [0, 1]$ , be a stationary Gaussian field with mean 0, variance 1, and covariance function  $\rho(\tau) = E[G(x)G(x + \tau)]$ . The first two moments of the lognormal translation field

$$A(x) = \frac{e^{G(x)} - e^{1/2}}{(e^2 - e)^{1/2}} \quad (6.62)$$

are  $E[A(x)] = 0$ ,  $E[A(x)^2] = 1$ , and  $\xi(\tau) = E[A(x)A(x + \tau)] = (1 - e^{\rho(\tau)})/(1 - e)$  ([12], Example 3.1), so that  $\xi(\tau) = \exp(-\alpha|\tau|)$  for  $\rho(\tau) = \ln(1 + (\exp(1) - 1)\exp(-\alpha|\tau|))$ ,  $\alpha > 0$ . The Gaussian field  $G(x)$  has continuous samples by the Kolmogorov condition ([40], Proposition 4.2, and Theorem 3.1 in this book) since  $E[(G(x + \tau) - G(x))^2] \leq (2(e - 1)/2)\tau^{1+\beta}$  for any  $\beta > 0$  and  $\tau \geq 0$ . Hence,  $A(x)$  has also continuous samples since the mapping  $G(x) \mapsto A(x)$  is continuous.

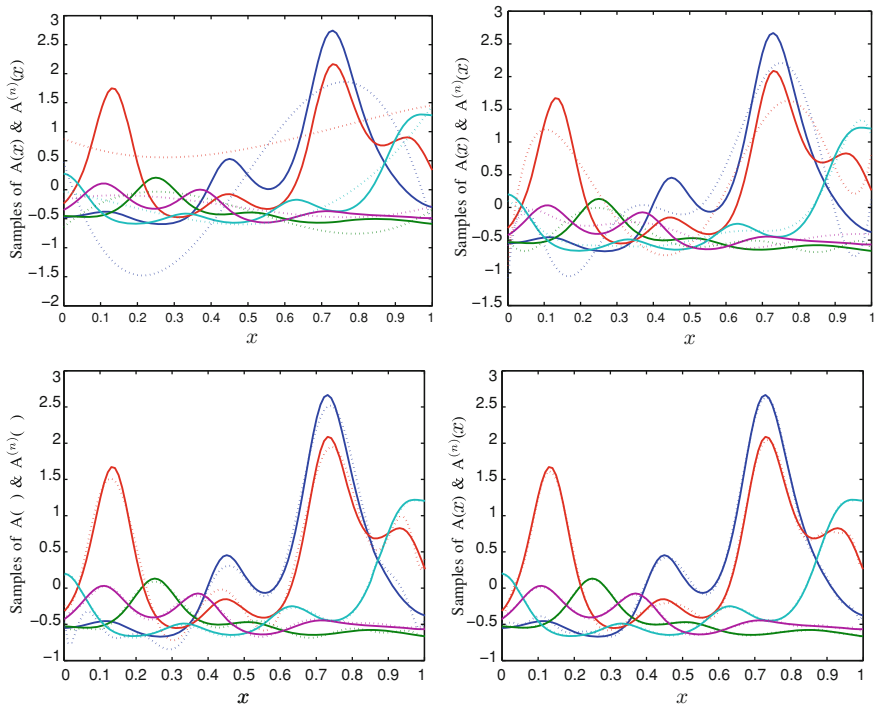
Consider the linear model in (6.40) with basis functions  $\{\theta_k(x)\}$  taken to be modified Chebyshev polynomials  $\{T_k(x)\}$  defined on  $[0, 1]$ , that is,

$$A^{(n)}(x) = \sum_{k=0}^n C_k T_k(x), \quad x \in D = [0, 1], \quad (6.63)$$

where  $k = 0$  is included in the summation for consistency with the indexing of Chebyshev polynomials. The Chebyshev polynomials satisfy the recurrence formula  $T_{k+1}(x) = 2(2x - 1)T_k(x) - T_{k-1}(x)$ ,  $k = 1, 2, \dots$ , with  $T_0(x) = 1$  and  $T_1(x) = 2x - 1$ , and can be obtained from the classical Chebyshev polynomials  $\tilde{T}_k(u) = \cos(k \arccos(u))$ ,  $k = 0, 1, \dots$ , defined on  $[-1, 1]$  by the change of variable  $u = 2x - 1$ , and the relationship  $\tilde{T}_{k+1}(u) = 2u\tilde{T}_k(u) - \tilde{T}_{k-1}(u)$ ,  $k = 1, 2, \dots$ , with  $\tilde{T}_0(u) = 1$  and  $\tilde{T}_1(u) = u$ . Note that the samples of  $A^{(n)}(x)$  belong to a subspace of  $C[0, 1]$  spanned by  $(T_0(x), T_1(x), \dots, T_n(x))$ , while the samples of  $A(x)$  are members of  $C[0, 1]$ .

Figure 6.14 shows with solid and dotted lines five samples of  $A(x)$  defined by (6.62) with  $\alpha = 10$  and the corresponding samples of  $A^{(n)}(x)$ , respectively. The samples of  $A^{(n)}(x)$  have been obtained by an optimization algorithm with objective function given by (6.48) for  $n = 4, 9, 14$ , and 19. The difference between the samples of  $A(x)$  and  $A^{(n)}(x)$  decreases with  $n$ , and becomes negligible at the figure scale for  $n = 19$ . Figure 6.15 shows with thin solid and dotted lines the real and the imaginary parts of estimates  $\hat{\varphi}_{A^{(n)}(x)}(u)$  of the characteristic function of  $A^{(n)}(x)$  in (6.57) with  $n = 9$  at  $x = 0.1, 0.4$ , and  $0.7$  based on  $n_s = 50$  (left panel) and  $n_s = 200$  (right panel) independent samples of  $(C_1, \dots, C_n)$  calculated by minimizing the objective function in (6.48). The real and the imaginary parts of the target characteristic function  $\varphi_A(u)$  are shown with heavy solid and dotted lines, respectively. The estimates of the characteristic function of  $A^{(n)}(x)$  are consistent with the target characteristic function and improve with the sample size.

The discrepancy between the target characteristic function  $\varphi_A(u)$  and its estimates  $\hat{\varphi}_{A^{(n)}(x)}(u)$  has a component related to model accuracy and a component related to statistical uncertainty corresponding to the first and the second terms of the bound in (6.53). The plots in Fig. 6.15 show that the statistical component of the discrepancy between  $\varphi_A(u)$  and  $\hat{\varphi}_{A^{(n)}(x)}(u)$  can be reduced by increasing the sample size  $n_s$ . The model component of this discrepancy depends on the accuracy of the representation  $A^{(n)}(x)$  of  $A(x)$ , and can be reduced by increasing the model order  $n$ .



**Fig. 6.14** Five samples of  $A(x)$  (solid lines) and  $A^{(n)}(x)$  (dotted lines) for  $n = 4$  (top left panel),  $n = 9$  (top right panel),  $n = 14$  (bottom left panel), and  $n = 19$  (bottom right panel)

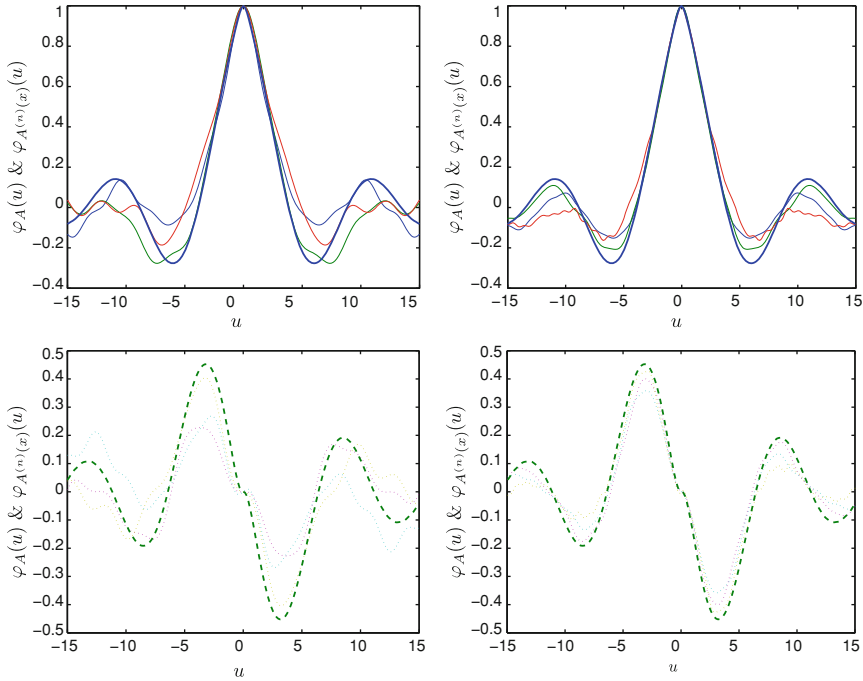
For example, the distance  $\sup_{-15 \leq u \leq 15} \{|\varphi_A(u) - \hat{\varphi}_{A^{(n)}(x)}(u)|\}$  between  $\varphi_A(u)$  and  $\hat{\varphi}_{A^{(n)}(x)}(u)$  is 0.2539, 0.2353, and 0.2904 at  $x = 0.1, 0.4$ , and  $0.7$  for  $n = 3$ , and decreases to 0.1332, 0.1137, and 0.1749 at  $x = 0.1, 0.4$ , and  $0.7$  for  $n = 9$ . The estimates  $\hat{\varphi}_{A^{(n)}(x)}(u)$  are based on  $n_s = 50$  (left panels) and  $n_s = 200$  (right panels) independent samples of the random coefficients  $\{C_k\}$  in the expression of  $A^{(n)}(x)$  given by (6.63).  $\diamond$

## 6.4 Exercises

**Exercise 6.1** Extend the Bayesian analysis in (6.5) to (6.8) to an  $\mathbb{R}^d$ -valued Gaussian variable with unknown second moment properties.

**Exercise 6.2** Construct plots as in Fig. 6.4 for sets of 10 and more independent samples of  $X$  and various prior densities  $f'(\sigma, \lambda, \rho)$ .

**Exercise 6.3** Extend considerations in Example 6.1 to an  $\mathbb{R}^d$ -valued translation vector with Beta distributed coordinates.



**Fig. 6.15** Heavy solid and dotted lines are real and imaginary parts of  $\varphi_A(u)$ . Thin solid and dotted lines are estimates of real and imaginary parts of  $\varphi_{A^{(n)}(x)}(u)$  at  $x = 0.1, 0.4$ , and  $0.7$  based on  $n_s = 50$  (left panels) and  $n_s = 200$  (right panels) samples

**Exercise 6.4** Develop estimates for extreme wind as in Fig. 6.6 using posterior densities for the uncertain parameters of the wind speed process, rather than point estimates.

**Exercise 6.5** Perform calculations as in Example 6.2 for a real-valued random field defined on a bounded rectangle in  $\mathbb{R}^d$ ,  $d \geq 2$ .

**Exercise 6.6** Propose alternatives to the models in (6.22) and (6.23) that are consistent with the available information.

**Exercise 6.7** Repeat the calculations in Example 6.3 for  $A_{c,k}^{(n)}$  and  $A_{s,k}^{(n)}$  having the same second moment properties as in this example but following different distributions.

**Exercise 6.8** Let  $X(t) = G(t)^3$ , where  $G(t) = A \cos(\nu t) + B \sin(\nu t)$ ,  $\nu > 0$ , and  $A$  and  $B$  are independent  $N(0, 1)$ . Show that  $X(t)$  has the representation in (6.44)

with uncorrelated but dependent random coefficients. Develop a linear model with dependent coefficients for  $X(t)$ .

**Exercise 6.9** Let  $G(x_1, x_2)$ ,  $(x_1, x_2) \in [0, 1]^2$ , be a real-valued Gaussian field defined by  $G(x_1, x_2) = (a_1^2 + a_2^2)^{-1/2}(a_1 G_1(x_1) + a_2 G_2(x_2))$ ,  $(x_1, x_2) \in [0, 1]^2$ , where  $a_1, a_2 \in \mathbb{R}$  and  $G_r(x_r)$ ,  $r = 1, 2$ , are two independent real-valued, stationary Gaussian random fields with mean 0, variance 1, and one-sided spectral densities  $g_r(\nu) = 1(0 \leq \nu \leq \bar{\nu}_r)/\bar{\nu}_r$ ,  $\bar{\nu}_r > 0$ . The image  $A(x_1, x_2)$  of  $G(x_1, x_2)$  given by (6.62) is a lognormal translation field defined in  $[0, 1]^2$ . Let

$$A^{(n)}(x_1, x_2) = \sum_{k,l=0}^n C_{kl} T_k(x_1) T_l(x_2), \quad (x_1, x_2) \in [0, 1]^2, \quad (6.64)$$

be a linear model for  $A(x_1, x_2)$ , where  $T_k(x_1)$  and  $T_l(x_2)$  are Chebyshev polynomials as in (6.63). Develop an optimization algorithm using the objective function in (6.48) to calculate samples of  $\{C_{kl}\}$  from samples of  $A(x_1, x_2)$ .

## References

1. Babuška IM, Liu K-M (2003) On solving stochastic initial-value differential equations. *Math Models Methods Appl Sci* 13(5):715–745
2. Babuška IM, Nobile F, Tempone R (2007) A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J Numer Anal* 45(3):1005–1034
3. Babuška IM, Tempone R, Zouraris E (2004) Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J Numer Anal* 42(2):800–825
4. Davis PJ (1975) Interpolation and approximation. Dover Publications, New York
5. Douglas JF, Garboczi EJ (1995) Intrinsic viscosity and polarizability of particles having a wide range of shapes. *Adv Chem Phys* 91:85–153
6. Duncan BS, Olson AJ (1999) Approximation and characterization of molecular surfaces. *J Comput Chem* 20:383–395
7. Elman HC, Ernst OG, O’Leary DP, Stewart M (2005) Efficient iterative algorithms for stochastic finite element method with applications to acoustic scattering. *Comput Methods Appl Mech Eng* 194:1037–1055
8. Field RV, Grigoriu M (2007) Model selection in applied science and engineering: A decision-theoretic approach. *J Eng Mech* 133(7):780–791
9. Foo J, Wan X, Karniadakis E (2008) The multi-element probabilistic collocation method (ME-PCM): Error analysis and applications. *J Comput Phys* 227:9572–9595
10. Frauenfelder P, Schwab C, Todor RA (2005) Finite element for elliptic problems with stochastic coefficients. *Comput Methods Appl Mech Eng* 194:205–228
11. Garboczi EJ (2002) Three-dimensional mathematical analysis of particle shape using X-ray tomography and spherical harmonics: Application to aggregates used in concrete. *Cem Concr Res* 32:1621–1638
12. Grigoriu M (1995) Applied non-Gaussian processes. Examples, theory, simulation, linear random vibration, and MATLAB solutions. Prentice Hall, Englewoods Cliffs
13. Grigoriu M (2006) Evaluation of Karhunen–Loève, spectral, and sampling representations for stochastic processes. *J Eng Mech ASCE* 132(2):179–189

14. Grigoriu M (2009) Algorithms for generating large sets of synthetic directional wind data for hurricane, thunderstorm, and synoptic winds. Technical note 1626. National Institute of Standard and Technology
15. Grigoriu M (2010) Probabilistic models for stochastic elliptic partial differential equations. *J Comput Phys* 229:8406–8429. doi:[10.1016/j.jcp.2010.07.023](https://doi.org/10.1016/j.jcp.2010.07.023)
16. Grigoriu M (2010) A spectral-based Monte Carlo algorithm for generating samples of nonstationary Gaussian processes. *Monte Carlo Methods Appl* 16(2):143–165. doi:[10.1515/MCMA.2010.006](https://doi.org/10.1515/MCMA.2010.006)
17. Grigoriu M, Field RV (2008) A solution of the static frame validation challenge problem using Bayesian model selection. *Computer methods in applied mechanics and engineering*, special issue on Sandia's validation challenge workshop
18. Grigoriu M, Garboczi EJ, Kafali C (2006) Spherical harmonic-based random fields for aggregate used in concrete. *Powder Technol* 166:123–138
19. Grigoriu M, Papoulia KD (2005) Effective conductivity by a probability-based local method. *J Appl Phys* 98:033706(1–10)
20. Hakobyan Y, Papoulia KD, Grigoriu MD (2007) Physical and geometrical percolation of effective conductivity on a lattice. *Phys Rev B* 76:144205
21. Hernández DB (1995) Lectures on probability and second order random fields. World Scientific, London
22. Hobson EW (1931) The theory of spherical and ellipsoidal harmonics. Cambridge University Press, Cambridge
23. Iranpour P, Chacon R (1988) The Mark Kac lectures. Macmillan, New York
24. Johnson N, Kotz S (1970) Distributions in statistics: Continuous univariate distributions-1. Houghton Mifflin Company, Boston
25. Johnson N, Kotz S (1970) Distributions in statistics: Continuous univariate distributions-2. Houghton Mifflin Company, Boston
26. Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, vol 2, 2nd edn. Wiley, New York
27. Larson HJ (1974) Introduction to the theory of statistics. Wiley, New York
28. Martins ES, Stedinger JR (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrological data. *Water Resour Res* 36(3):737–744
29. Max NL, Getzoff ED (1988) Spherical harmonic molecular surfaces. *IEEE Comput Graph Appl* 8:42–50
30. Powell MJD (1981) Approximation theory and methods. Cambridge University Press, Cambridge
31. Raiffa H, Schlaifer R (1961) Applied statistical decision theory. MIT Press, Cambridge
32. Rektorys K (1969) Survey of applicable mathematics. MIT Press, Cambridge
33. Resnick SI (1998) A probability path. Birkhäuser, Boston
34. Ritchie DW, Kemp GJL (1999) Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces. *J Comput Chem* 20(4):383–395
35. Schwab C, Todor RA (2006) Karhunen–Loève approximation of random fields by generalized fast multipole methods. *J Comput Phys* 217:100–122
36. Simiu E, Scanlan RH (1986) Wind effects on structures. An introduction to wind engineering, 2nd edn. Wiley, New York
37. Simiu E, Scanlan R (1996) Wind effects on structures: Fundamentals and applications to design. Wiley, New York
38. Tolstov GP (1962) Fourier series. Dover Publications, New York
39. Torquato S (2002) Random heterogeneous materials microstructure and macroscopic properties. Springer, New York
40. Wong E, Hajek B (1985) Stochastic processes in engineering systems. Springer, New York
41. Zellner A (1971) An introduction to Bayesian inference in econometrics. Wiley, New York

# Chapter 7

## Stochastic Ordinary Differential and Difference Equations

### 7.1 Introduction

Differential and difference equations with deterministic and/or random coefficients and input are used extensively in applications to describe the behavior of a broad range of physical systems. It is common to describe the states of dynamic systems in random environment by solutions of ordinary differential or difference equations with random input and deterministic or random coefficients depending on the uncertainty in the system properties. We refer to equations with deterministic coefficients and random input as stochastic equations, and equations with both random input and coefficients as stochastic equations with random coefficients. Differential and difference equations with random coefficients and input can be divided in three classes: (i) equations with state independent, time invariant random coefficients, (ii) equations with state independent, time dependent random coefficients, and (iii) equations with state dependent random coefficients.

Equations with state independent, time invariant random coefficients may result from stochastic partial differential equations by spatial discretization via finite difference, finite element, or other approximate methods [1–3]. The finite difference representation (1.5) of the partial differential transport equation (1.3) is an ordinary differential equation of this type. Depending on whether time argument is or not discretized, the resulting equations are differential or difference equations with random coefficients. A summary of methods for analyzing this class of stochastic equations can be found in [4] (Sects. 8.4.1, 8.8, and 9.2).

Equations with state independent, time variant random coefficients are encountered in economics, sociology, engineering, and other applied fields, and describe situations in which modeling conditions change over time. For example, the amplitude of a mathematical pendulum with length varying randomly in time according to a specified law satisfies a second order ordinary differential equation with random coefficients [5]. Differential equations with white noise coefficients, referred to in random vibration theory as equations with multiplicative noise, are used to establish conditions for the stability of the state of dynamic systems in random environment

([4], Sec. 8.7, [6]). Examples of equations of this type are in (1.7) and (1.9). Stability conditions and moment equations have also been developed for solutions of differential/difference equations with random coefficients evolving in time according to semi-Markov processes [7, 8, 9, 10, 11, 12]. Difference equations with state independent, time dependent random coefficients are used extensively in economics, sociology, and stochastic hydrology [13, 14], and in a broad range of applications in mechanics [7, 8, 9, 10, 11, 12]. This class of equations has also been employed to construct non-Gaussian models ([15], Sect. 4.1). Some of the non-Gaussian models have specified marginal distribution and correlation functions ([16] and [17], Sect. 3.6.1).

Equations with state dependent random coefficients are relevant in many applications, for example, they have been used to characterize damage evolution in Daniels systems [18, 19] and systems with initial cracks ([20], Sect. 7.5.2.3) subjected to Gaussian actions, and calculate the reliability of ideal elasto-plastic oscillators subjected to Gaussian white noise [21]. Difference equations with state dependent random coefficients, for example, GARCH and ARCH models ([22], Chap. 10) and threshold autoregressive models ([15], Sect. 4.2) have been used in finance to capture market volatility and characterize river flows, respectively. The last section of this chapter will present examples of differential equations of this type.

A common feature of the solutions of differential/difference equations with random coefficients is that they are non-Gaussian processes/time series. The probability laws of these solutions are determined by the functional form of their defining equations and the properties of both the random coefficients of and the input to these equations. Equations with state independent, time invariant random coefficients and with state dependent random coefficients are the simplest and most difficult to solve, respectively. The analysis of the latter type of equations poses notable difficulties since their coefficients are functionals of state history.

Monte Carlo simulation is the only general method for solving equations with random coefficients. Analytical solutions for these equations have been obtained in special cases that, generally, are of limited practical interest. A host of approximate methods has been proposed for solving equations with random coefficients under the assumption that the uncertainty in their coefficients can be captured by a finite number of random variables.

Let  $X(t)$  be an  $\mathbb{R}^d$ -valued stochastic process satisfying the differential equation

$$dX(t) = a(X(t-), Y(t-)) dt + b(X(t-), Y(t-)) dS(t), \quad t \geq 0, \quad (7.1)$$

where the entries of the  $(d, 1)$  and  $(d, d')$  matrices  $a$  and  $b$  are real-valued functions,  $S$  denotes an  $d'$ -dimensional semimartingale, and  $Y$  is an  $\mathbb{R}^{d_Y}$ -valued stochastic process with specified probability law. The processes  $S$  and  $Y$  are assumed to be mutually independent and independent of the initial state  $X(0)$ .

In the special case in which  $Y(t)$  is time invariant, that is, an  $\mathbb{R}^{d_Y}$ -valued random variable denoted by  $\Theta$ , (7.1) becomes

$$dX(t) = a(X(t-), \Theta) dt + b(X(t-), \Theta) dS(t), \quad t \geq 0. \quad (7.2)$$

Note that (7.1) and (7.2) are stochastic differential equations of the type studied in Sect. 5.5.1 on samples of  $Y(t)$  and  $\Theta$ , respectively.



**Definition 7.1** The stochastic equations in (7.1) and (7.2) are referred to as equations with state independent, time variant and time invariant random coefficients, respectively. If the coefficients  $a$  and/or  $b$  in (7.1) and (7.2) are functions of the entire state history rather than its current value, in addition to  $Y$  or  $\Theta$ , then (7.1) and (7.2) are called stochastic equations with state dependent random coefficients.

Our objectives are to establish conditions under which stochastic equations with deterministic and random coefficients have solutions that are unique, and explore methods for solving these equations. Heuristic and rigorous methods for solving stochastic equations are discussed. Both continuous and discrete time stochastic equations are considered. Section 7.2 examines stochastic equations with deterministic coefficients of the type considered in random vibration. Discrete and continuous time systems are discussed in Sects. 7.2.1, 7.2.2 and 7.2.3. Developments in continuous time systems are based on Itô's formula rather than heuristic arguments. Methods for solving discrete and continuous time systems with random coefficients and input are discussed in Sects. 7.3 and 7.4. The methods include Monte Carlo, conditional simulation, state augmentation, stochastic reduced order models, stochastic Galerkin, stochastic collocation, and simplified techniques for the case in which the uncertainty in the random coefficients of these equations is small. Theoretical developments in this chapter are applied in Sect. 7.5 to study stochastic stability and noise induced transitions for some dynamic systems, discuss a class of random vibration problems of practical interest, and study the behavior of a simple degrading system whose state satisfies a stochastic differential equation with state dependent coefficients.

## 7.2 Stochastic Equations with Deterministic Coefficients

Equations of the type as in (7.2) with coefficients that do not depend on  $\Theta$  are considered. The driving noise  $S$  can be Gaussian, Poisson, Lévy, or any semimartingale. If the drift depends linearly on the state and the diffusion is state independent, the resulting equation defines a linear random vibration problem. Otherwise, we deal with a nonlinear random vibration problem.

The presentation of essentials on discrete time linear systems is followed by heuristic and rigorous methods for solving linear and nonlinear random vibration problems.

### 7.2.1 Discrete Time Linear Systems

Let  $X(n)$ ,  $n = 0, 1, \dots$ , be an  $\mathbb{R}^d$ -valued time series defined by the recurrence formula

$$X(n+1) = a(n)X(n) + b(n)W(n), \dots n = 0, 1, \dots, \quad (7.3)$$

where  $a(n)$  and  $b(n)$  denote  $(d, d)$  and  $(d, d')$  deterministic matrices, and  $W(n)$ ,  $n = 0, 1, \dots$ , is an  $\mathbb{R}^{d'}$ -valued, weakly stationary, and uncorrelated time series with mean vector  $\mu_w(n) = E[W(n)]$  and covariance matrix  $\gamma_w(n) = E[(W(n) - \mu_w)(W(n) - \mu_w)']$ . It is assumed that initial state  $X(0)$  and the driving noise  $\{W(n)\}$  are uncorrelated.

We develop recurrence formulas for the mean  $\mu(n) = E[X(n)]$ , covariance  $\gamma(n) = E[(X(n) - \mu(n))(X(n) - \mu(n))']$ , and covariance function  $c(m, n) = E[(X(m) - \mu(m))(X(n) - \mu(n))']$  of the system state. It is assumed that the mean  $\mu_0 = \mu(0) = E[X(0)]$  and the covariance  $\gamma_0 = \gamma(0) = E[(X(0) - \mu_0)(X(0) - \mu_0)']$  of  $X(0)$ , the second moment properties of  $\{W(n)\}$ , and the matrices  $a(n)$  and  $b(n)$  are known.

**Theorem 7.1** *The mean, covariance, and covariance function of the state vector defined by (7.3) satisfy the equations*

$$\begin{aligned}\mu(n+1) &= a(n)\mu(n) + b(n)\mu_w(n), \\ \gamma(n+1) &= a(n)\gamma(n)a(n)' + b(n)\gamma_w(n)b(n)', \quad \text{and} \\ c(m, n) &= \alpha(n, m)\gamma(m), \quad n > m\end{aligned}\tag{7.4}$$

with initial conditions  $\mu(0) = \mu_0$ ,  $\gamma(0) = \gamma_0$ , and  $c(m, m) = \gamma(m)$ , where  $\alpha(n, m)$  denotes the state transition matrix.

*Proof* The mean equation results by averaging (7.3). Since  $\tilde{X}(n) = X(n) - \mu(n)$  satisfies the finite difference equation  $\tilde{X}(n+1) = a(n)\tilde{X}(n) + b(n)\tilde{W}(n)$  with  $\tilde{W}(n) = W(n) - \mu_w(n)$ , the second equation in (7.4) results from the recurrence formula for  $\tilde{X}$  by direct calculations using the absence of correlation between  $X(n)$  and  $W(n)$ . For the covariance function,  $c(m, n) = E[\tilde{X}(n)\tilde{X}(m)']$  with

$$\tilde{X}(n) = \alpha(n, m)\tilde{X}(m) + \sum_{k=m}^{n-1} \alpha(n, k+1)b(k)\tilde{W}(k), \quad n > m,$$

gives the third formula in (7.4), where  $\alpha(n, m) = \prod_{k=m}^{n-1} a(k)$  with the convention that  $\alpha(n, n) = I$  is the  $(d, d)$ -identity matrix.  $\blacktriangle$

Consider the special case of a time invariant system driven by a weakly stationary noise, that is,  $a(n) = a$ ,  $b(n) = b$ ,  $\mu_w(n) = \mu_w$ , and  $\gamma_w(n) = \gamma_w$  in (7.3). Under some conditions,  $X(n)$  becomes a weakly stationary process as  $n \rightarrow \infty$ . For example, the asymptotic mean  $\mu = \lim_{n \rightarrow \infty} \mu(n)$  of  $X(n)$  can be calculated from  $\mu = (I - a)^{-1}b\mu_w$  provided that all eigenvalues of  $a$  are included in the open ball with unit radius centered at the origin of  $\mathbb{C}$ . This follows from the first equality in (7.4) whose asymptotic form is  $\mu = a\mu + b\mu_w$ . Similar results can be obtained for the other moments of the state.

*Example 7.1* Suppose  $X(n)$  satisfies (7.3) with  $d = 2$ ,  $d' = 1$ ,  $a_{11} = \beta_1 \neq 0$ ,  $a_{22} = \beta_2 \neq 0$ ,  $a_{12} = a_{21} = 0$ , and  $b(n)$  is the unit vector. If  $|\beta_k| < 1$ ,  $k = 1, 2$ , then  $X(n)$  becomes weakly stationary as  $n \rightarrow \infty$  with means  $\mu_k = \lim_{n \rightarrow \infty} E[X_k(n)] =$

$\mu_w/(1 - \beta_k)$ ,  $k = 1, 2$ , and covariances  $\gamma_{kl} = \lim_{n \rightarrow \infty} E[\tilde{X}_k(n)\tilde{X}_l(n)] = \gamma_w/(1 - \beta_k\beta_l)$ ,  $k = 1, 2$ .  $\diamond$

*Proof* The mean equation in (7.4) gives  $\mu_k = \beta_k\mu_k + \mu_w$  as  $n \rightarrow \infty$ , so that  $\mu_k = \mu_w/(1 - \beta_k)$  under the stated assumptions. Similar calculations using the second equality in (7.4) give stationary state covariances.  $\blacktriangle$

Note that the partial, second moment characterization of  $W(n)$  in (7.3) is insufficient to characterize  $X(n)$  beyond its second moment properties. The probability law of  $\{W(n)\}$  needs to be specified for this purpose. For example, if the random variables  $\{W(n)\}$  are independent with characteristic function  $\varphi_{W_n}(u) = E[\exp(iuW(n))]$ , the characteristic function  $\varphi_n(u) = E[\exp(iuX(n))]$  of  $X(n)$  in (7.3) with  $d=1$  satisfies the recurrence formula

$$\varphi_{n+1}(u) = E[e^{iu(a(n)X(n)+b(n)W(n))}] = \varphi_n(a(n)u)\varphi_{W_n}(b(n)u), \quad u \in \mathbb{R}.$$

The formula can be applied to calculate the characteristic function of  $X(n)$ ,  $n \geq 1$ , starting from the characteristic function of the initial state  $X(0)$ . Similar arguments can be used to develop recurrence formulas for joint characteristic functions. For example, the characteristic function  $\varphi_{n,n+2}(u, v) = E[\exp(i(uX(n) + vX(n+2)))]$  of  $(X(n), X(n+2))$  can be calculated from

$$\varphi_{n,n+2}(u, v) = \varphi_n(u + a(n)a(n+1)v)\varphi_{W(n)}(a(n+1)b(n)v)\varphi_{W(n+1)}(b(n+1)v)$$

since  $X(n+2) = a(n+1)a(n)X(n) + a(n+1)b(n)W(n) + b(n+1)W(n+1)$  and  $X(n)$ ,  $W(n)$ , and  $W(n+1)$  are independent random variables.

### 7.2.2 Continuous Time Linear Systems

The continuous time analog of (7.3) is

$$\dot{X}(t) = a(t)X(t) + b(t)W(t), \quad t \geq 0, \quad (7.5)$$

where  $X(t)$  is an  $\mathbb{R}^d$ -valued stochastic process,  $W(t)$  denotes an  $\mathbb{R}^{d'}$ -valued white noise process with mean  $\mu_w(t)$  and covariance function  $E[(W(s) - \mu_w(s))(W(t) - \mu_w(t))'] = \gamma_w(t)\delta(s - t)$ ,  $\gamma_w(t)$  is a positive definite  $(d', d')$  matrix, and  $a(t)$  and  $b(t)$  are  $(d, d)$  and  $(d, d')$  matrices. The initial state  $X(0)$  has mean  $\mu_0$  and covariance  $\gamma_0$ , and is uncorrelated from the driving noise.

The classical linear random vibration theory provides equations for the mean  $\mu(t) = E[X(t)]$ , covariance  $\gamma(t) = E[\tilde{X}(t)\tilde{X}(t)']$ , and covariance function  $c(s, t) = E[\tilde{X}(t)\tilde{X}(s)']$  of  $X(t)$  defined by (7.5), where  $\tilde{X}(t) = X(t) - \mu(t)$ . The development of these equations is based on heuristic arguments, calculations, and definitions of white noise. For example, note that  $X(t)$  in (7.5) is not defined since  $W(t)$  does not exist in the second moment sense. Formal calculations used in the classical theory of linear random vibration give

$$\begin{aligned}
\dot{\mu}(t) &= a(t)\mu(t) + b(t)\mu_w(t), \\
\dot{\gamma}(t) &= a(t)\gamma(t) + \gamma(t)b(t)a(t)' + b(t)\gamma_w(t)b(t)', \quad \text{and} \\
\frac{\partial c(s, t)}{\partial t} &= a(t)c(s, t), \quad t > s,
\end{aligned} \tag{7.6}$$

with initial conditions  $\mu(0) = \mu_0$ ,  $\gamma(0) = \gamma_0$ , and  $c(s, s) = \gamma(s)$  ([23], Sect. 6.2, and [20], Sect. 5.2.1). The derivation of these equations uses the representation

$$X(t) = \alpha(t, 0)X(0) + \int_0^t \alpha(t, s)b(s)W(s)ds, \quad t \geq 0, \tag{7.7}$$

where system transition matrix  $\alpha(t, s)$ ,  $t > s$ , is the solution of

$$\frac{\partial \alpha(t, s)}{\partial t} = a(t)\alpha(t, s), \quad t > s, \tag{7.8}$$

with  $\alpha(s, s) = I$  ([24], Sect. 1.3 and Theorem 1, p. 40).

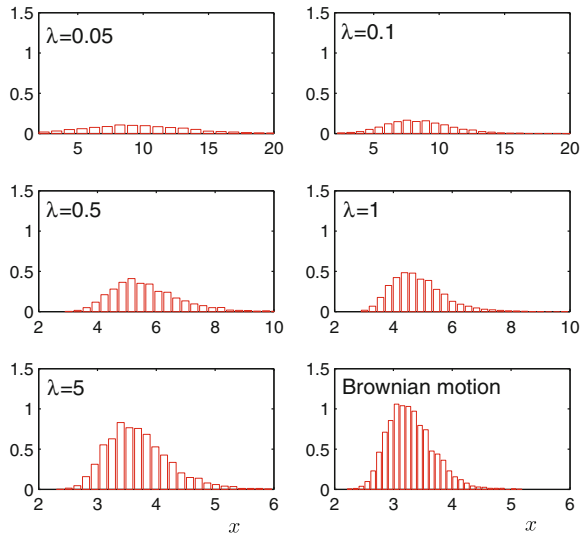
*Example 7.2* Let  $X(t)$ ,  $t \geq 0$ , be the solution of (7.5) with  $d = d' = 1$ ,  $a(t) = -\rho$ ,  $\rho > 0$ ,  $b(t) = 1$ ,  $\gamma_w(t) = \gamma_w$ , and  $\mu_0 = 0$ . The differential equations  $\dot{\mu}(t) = -\rho\mu(t)$ ,  $\dot{\gamma}(t) = -2\rho\gamma(t) + \gamma_w$ , and  $\partial c(s, t)/\partial t = -\rho c(s, t)$  with  $\mu(0) = 0$ ,  $\gamma(0) = \gamma_0$ , and  $c(s, s) = \gamma(s)$  given by (7.6) have the solutions  $\mu(t) = 0$ ,  $\gamma(t) = (\gamma_0 - \gamma_w/(2\rho))\exp(-2\rho t) + \gamma_w/(2\rho)$ , and  $c(s, t) = \gamma(s)\exp(-\rho(t - s))$ ,  $t > s$ . Additional examples can be found in [20] (Chap. 5).  $\diamond$

The heuristic approach of classical linear random vibration theory delivers differential equations for the second moment properties of  $X(t)$ , but cannot be extended to find higher order properties of this process. This is a significant limitation since higher order statistics of  $X(t)$  are sensitive to noise type. For example, let  $X(t)$  be the solution of (7.5) with  $d = d' = 1$ ,  $a(t) = -\rho$ ,  $\rho > 0$ ,  $b(t) = \sqrt{2\rho}$ , and suppose  $W(t)$  is a Gaussian or a Poisson white noise, so that  $X(t)$  is the solution of

$$\begin{aligned}
dX(t) &= -\rho X(t) dt + \sqrt{2\rho} dB(t) \quad \text{or} \\
dX(t) &= -\rho X(t) dt + \sqrt{2\rho} dC(t),
\end{aligned} \tag{7.9}$$

where  $B(t)$  is a Brownian motion,  $C(t) = \sum_{k=1}^{N(t)} Y_k$  denotes a compound Poisson process,  $N(t)$  is a homogeneous Poisson process with intensity  $\lambda > 0$ , and  $\{Y_k\}$  are iid random variables with finite variance. If  $E[Y_1] = 0$  and  $\lambda E[Y_1^2] = 1$ , then  $B(t)$  and  $C(t)$  are equal in the second moment sense so that the processes  $X(t)$  in (7.9) also have the same second moment properties. Yet, their higher order properties differ significantly. Figure 7.1 shows histograms of  $\max_{0 \leq t \leq 100} \{X(t)\}$  for  $\rho = 1$  constructed from 5000 independent samples of  $X(t)$ . Although the processes  $X(t)$  in this figures are equal in the second moment sense, the histograms of  $\max_{0 \leq t \leq 100} \{X(t)\}$  under Poisson white noise for various value of  $\lambda$  and under Gaussian white noise differ significantly. The histograms of  $\max_{0 \leq t \leq 100} \{X(t)\}$  under Poisson white noise approach the histogram of  $\max_{0 \leq t \leq 100} \{X(t)\}$  under Gaussian white noise as  $\lambda$  increases since  $C(t)$  with  $\lambda E[Y_1^2] = 1$  converges weakly to  $B(t)$  as  $\lambda \rightarrow \infty$  (Theorem 7.23).

**Fig. 7.1** Histograms of  $\max_{0 \leq s \leq 100} \{X_n(s)\}$  and  $\max_{0 \leq s \leq 100} \{X(s)\}$  for  $\rho = 1$



The remainder of this section reformulates the linear random vibration problem in (7.5) by defining white noise processes as formal derivatives of Brownian or other processes and using Itô's formula to construct differential equations for moments and other properties of  $X(t)$ . Linear random vibration problems are defined by

$$dX(t) = a(t)X(t)dt + b(t)dS(t), \quad t \geq 0, \quad (7.10)$$

where  $S(t)$  can be an  $\mathbb{R}^{d'}$ -valued semimartingale. As previously, the initial state  $X(0)$  is assumed to be independent of the noise. Our focus is on Gaussian and Poisson white noise processes as in (7.9), that is,  $S(t)$  is either  $B(t)$  or  $C(t)$ . If  $S(t) = B(t)$ , then  $X(t)$  is a Gaussian process, so that its probability law is completely defined by its second moment properties.

**Theorem 7.2** *If  $S(t) = B(t)$  in (7.10) is an  $\mathbb{R}^{d'}$ -valued Brownian motion with independent coordinates, the mean vector  $\mu(t) = E[X(t)]$ , covariance matrix  $\gamma(t) = E[\tilde{X}(t)\tilde{X}(t)']$ , and covariance function  $c(s, t) = E[\tilde{X}(t)\tilde{X}(s)']$ ,  $t > s$ , of the solution  $X(t)$  of this equation can be obtained from (7.6) with  $b(t)b(t)'$  in place of  $b(t)\gamma_w(t)b(t)'$ , where  $\tilde{X}(t) = X(t) - \mu(t)$ .*

*Proof* Itô's formula for  $\mathbb{R}^d$ -valued continuous semimartingales in Theorem 5.3 applied to the mappings  $x \mapsto g(x) = x_p$  and  $x \mapsto g(x) = x_px_q$  gives by averaging equations for the mean and the covariance functions of  $X(t)$ , where  $x_p$  and  $x_q$  are coordinates of  $x \in \mathbb{R}^d$ . For example, this formula applied to  $g(x) = x_px_q$  gives

$$\begin{aligned}
X_p(t)X_q(t) - X_p(0)X_q(0) &= \sum_{i=1}^d \int_0^t (\delta_{ip}X_q(s) + X_p(s)\delta_{iq})dX_i(s) \\
&+ \frac{1}{2} \sum_{i,j=1}^d \sum_{k=1}^{d'} \int_0^t (\delta_{ip}\delta_{jq} + \delta_{jp}\delta_{iq})b_{ik}(s)b_{jk}(s)ds \\
&= \int_0^t \left[ X_q(s) \sum_{j=1}^d a_{pj}(s)X_j(s) + X_p(s) \sum_{j=1}^d a_{qj}(s)X_j(s) \right] ds \\
&+ \frac{1}{2} \int_0^t [(b(s)b(s'))_{pq} + (b(s)b(s'))_{qp}]ds + \text{martingale}(t).
\end{aligned}$$

The expectation of the left side of this equation is  $r_{pq}(t) - r_{pq}(0)$ , where  $r(t) = E[X(t)X(t)']$  denotes the correlation matrix of  $X(t)$ . The right side has three terms. The terms of the form  $\int_0^t X_q(s) dB_r(s)$  have zero expectation. The order of integration can be changed in the first two terms on the right side of the equation by Fubini's theorem since  $X(t)$  is measurable with respect to both arguments and has finite expectation. For example, the first term becomes  $\int_0^t \sum_{j=1}^d a_{pj}(s)r_{qj}(s)ds$ . Differentiation with respect to time gives

$$\begin{aligned}
\dot{r}_{pq}(t) &= \sum_{j=1}^d a_{pj}(t)r_{qj}(t) + \sum_{j=1}^d a_{qj}(t)r_{pj}(t) + \frac{1}{2} \left[ (b(t)b(t'))_{pq} + (b(t)b(t'))_{qp} \right] \\
&= (a(t)r(t))_{pq} + (r(t)a(t'))_{pq} + (b(t)b(t'))_{pq}, \quad p, q = 1, \dots, d,
\end{aligned}$$

which yields the second equality in (7.6) for the case in which  $\gamma_w(t)$  is the identity matrix. The matrix  $\gamma_w(t)$  is identity since  $B(t)$  has independent coordinates.

The last equation in (7.6) results from the expectation of the integral form  $\tilde{X}(t) - \tilde{X}(s) = \int_s^t a(u)\tilde{X}(u)du + \int_s^t b(u)dB(u)$ ,  $t > s$ , of (7.10) following multiplication with  $\tilde{X}(s)$ , expectation, and differentiation with respect to time. We have

$$E[(\tilde{X}(t) - \tilde{X}(s))\tilde{X}(s)'] = E\left[\left(\int_s^t a(u)\tilde{X}(u)du + \int_s^t b(u)dB(u)\right)\tilde{X}(s)'\right].$$

The left side of this equation is  $c(t, s) - c(s, s)$ . The right side is  $\int_s^t a(u)c(u, s)du$  by Fubini's theorem, the independence between the random variables  $\int_s^t b(u)dB(u)$  and  $\tilde{X}(s)$ , and  $E[\int_s^t b(u)dB(u)] = 0$ . The differentiation of the resulting equation gives the last formula in (7.6).  $\blacktriangle$

Arguments used to prove Theorem 7.2 can be applied to derive equations for moments of order three and higher, characteristic functions, and other descriptors of  $X(t)$ . These equations are only needed for systems driven by non-Gaussian noise since, generally, their states are non-Gaussian processes.

*Example 7.3* Let  $X(t)$ ,  $t \geq 0$ , satisfy the first equation in (7.9) with  $\rho > 0$  and  $Y(t)^k dt$  in place of  $dB(t)$ , where  $k \geq 1$  is an integer and  $Y(t)$  is defined

by  $dY(t) = -\alpha Y(t)dt + \sqrt{2\alpha}dB(t)$ ,  $t \geq 0$ , with  $\alpha > 0$ . The moments  $\mu(p, q; t) = E[X(t)^p Y(t)^q]$  of order  $p + q$ ,  $p, q \geq 0$ , of  $(X(t), Y(t))$  satisfy the differential equation

$$\begin{aligned}\dot{\mu}(p, q; t) = & -p\rho\mu(p, q; t) + p\mu(p-1, q+k; t) \\ & -q\alpha\mu(p, q; t) + \alpha q(q-1)\mu(p, q-2; t)\end{aligned}\quad (7.11)$$

with the convention  $\mu(p, q; t) = 0$  if  $p$  and/or  $q$  are negative. The initial condition  $\mu(p, q; 0)$  results from the properties of  $(X(0), Y(0))$ , that need to be specified.

The moment equations in (7.11) are closed so that they can be solved exactly. Suppose that our objective is to solve moment equations with  $p + q = m$  and that moment equations with  $p + q < m$  have been solved. There are  $m + 1$  moment equations with  $p + q = m \geq 1$  involving, in addition to  $m + 1$  moments of the type  $\mu(p, q; t)$ , the moments  $\mu(p-1, q+k; t)$  and  $\mu(p, q-2; t)$  which need to be obtained prior to solving these moment equations. The moments  $\mu(p, q-2; t)$  are available since they are solutions of moment equations with  $p + q < m$ . The moments  $\mu(p-1, q+k; t)$  are unknown, but can be obtained from (7.11). The moment equation for  $p = 1$  and  $q = r$  is  $\dot{\mu}(1, r; t) = -\rho\mu(1, r; t) + \mu(0, r+k; t) - \alpha r\mu(1, r; t) + \alpha r(r-1)\mu(1, r-2; t)$ , and can be solved sequentially for increasing  $r$  using the fact that  $\mu(0, r+k; t)$  is known since  $Y(t)$  is Gaussian. These equations give  $\mu(1, r; t)$  for any  $r \geq 1$ . Similar calculations for  $p = 2$  and increasing  $r$  deliver  $\mu(2, r; t)$ , and so on.

For example, there are two moment equations for  $m = 1$  involving the moments  $\mu(1, 0; t)$ ,  $\mu(0, 1; t)$ , and  $\mu(0, k; t)$ . Since  $\mu(0, k; t)$  can be calculated from the defining equations for  $Y(t)$ , it is possible to find  $\mu(1, 0; t)$  and  $\mu(0, 1; t)$ . There are three moment equations for  $m=2$  involving the moments  $\mu(2, 0; t)$ ,  $\mu(1, 1; t)$ ,  $\mu(0, 2; t)$ ,  $\mu(1, k; t)$ , and  $\mu(0, k+1; t)$ . To solve these equations, we need only  $\mu(1, k; t)$  since  $\mu(0, k+1; t)$  is available. The sequence of equations for  $p = 1$  and increasing  $q = r$  mentioned previously can be used to find  $\mu(1, k; t)$ . Additional examples of linear systems driven by polynomials of filtered Gaussian and Poisson processes can be found in [4] (Examples 7.23, 7.26, and 7.28).  $\diamond$

*Proof* Note that  $(X(t), Y(t))$  is a bivariate diffusion process defined by the stochastic differential equation

$$d \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \begin{bmatrix} -\rho X(t) + Y(t)^k \\ -\alpha Y(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ \sqrt{2\alpha} \end{bmatrix} dB(t).$$

The integral form of Itô's formula in (5.15) applied to the mapping  $(X(t), Y(t)) \mapsto X(t)^p Y(t)^q$  gives

$$\begin{aligned}X(t)^p Y(t)^q - X(0)^p Y(0)^q = & \int_0^t p X(s)^{p-1} Y(s)^q dX(s) + \int_0^t q X(s)^p Y(s)^{q-1} dY(s) \\ & + \frac{1}{2} \int_0^t q(q-1) X(s)^p Y(s)^{q-2} (2\alpha) ds.\end{aligned}$$

The expectation of the left side of this equation is  $\mu(p, q; t) - \mu(p, q; 0)$ . The first and second integrals on the right side of the equation are  $p \int_0^t (-\rho X(s)^p Y(s)^q + X(s)^{p-1} Y(s)^{q+k}) ds$  and  $\int_0^t q X(s)^p Y(s)^{q-1} (-\alpha Y(s) ds + \sqrt{2\alpha} dB(s))$  by the definition of  $(X(t), Y(t))$ . Since  $(X(t), Y(t))$  is measurable in both arguments and integrable, the expectation of, for example, the first integral is  $p \int_0^t (-\rho \mu(p, q; s) + \mu(p-1, q+k; s)) ds$  by Fubini's theorem. Similar arguments show that the expectation of the above expression is

$$\begin{aligned} \mu(p, q; t) - \mu(p, q; 0) &= p \int_0^t (-\rho \mu(p, q; s) + \mu(p-1, q+k; s)) ds \\ &\quad - \alpha q \int_0^t \mu(p, q; s) ds + \frac{q(q-1)}{2} (2\alpha) \int_0^t \mu(p, q-2; s) ds. \end{aligned}$$

Differentiation with respect to time  $t$  yields (7.11).  $\blacktriangle$

The formulas in (7.6) cannot be used to find the second moment properties for the stochastic process in Example 7.3 since the driving noise is not white. Two options are available. We can extend the formulas in (7.6) such that they apply to colored noise ([20], Sect. 5.2.1) or perform direct calculations by using properties of Gaussian variables and/or multiple Wiener-Itô integrals. These options are illustrated by the following example.

*Example 7.4* The stationary covariance function of  $X(t)$  in Example 7.3 with  $k = 2$  has the expression

$$c(t, s) = \frac{2}{\rho(\rho + 2\alpha)} e^{-\rho(t-s)} + \frac{2}{\rho^2 - 4\alpha^2} (e^{-2\alpha(t-s)} - e^{-\rho(t-s)}), \quad t > s, \quad (7.12)$$

and can be obtained by an extension of the mean and covariance equations in (7.6) to the case in which the driving noise is colored or by direct calculations.  $\diamond$

*Proof* Consider first an extension of (7.7). Let  $X(t)$  be defined by (7.5) with  $Z(t)$  in place of  $W(t)$ , where  $Z(t)$  is an  $\mathbb{R}^{d'}$ -valued colored noise whose coordinates have finite variance, so that it can be calculated from (7.7) with  $W(t)$  replaced by  $Z(t)$ . The centered process  $\tilde{X}(t) = X(t) - E[X(t)]$  satisfies the equation

$$\tilde{X}(t) = \alpha(t, 0) \tilde{X}(0) + \int_0^t \alpha(t, s) b(s) \tilde{Z}(s) ds, \quad t \geq 0, \quad (7.13)$$

where  $\tilde{Z}(t) = Z(t) - E[Z(t)]$ . Let  $\gamma_0 = E[\tilde{X}(0) \tilde{X}(0)']$  and  $c_z(u, v) = E[\tilde{Z}(u) \tilde{Z}(v)']$  denote the covariance of the initial state and the covariance function of  $Z$ . For  $t > s$ , we have



$$\begin{aligned}
\tilde{X}(t)\tilde{X}(s)' &= \left( \alpha(t, 0)\tilde{X}(0) + \int_0^t \alpha(t, u)b(u)\tilde{Z}(u)du \right) \\
&\quad \times \left( \alpha(s, 0)\tilde{X}(0) + \int_0^s \alpha(s, v)b(v)\tilde{Z}(v)dv \right)' \\
&= \alpha(t, 0)\tilde{X}(0)\tilde{X}(0)'\alpha(s, 0)' + \alpha(t, 0)\tilde{X}(0) \int_0^s \tilde{Z}(v)'b(v)'\alpha(s, v)'dv \\
&\quad + \left( \int_0^t \alpha(t, u)b(u)\tilde{Z}(u)du \right) \tilde{X}(0)'\alpha(s, 0)' \\
&\quad + \int_0^t \int_0^s \alpha(t, u)b(u)\tilde{Z}(u)\tilde{Z}(v)'b(v)'\alpha(s, v)'dudv,
\end{aligned}$$

which gives

$$c(t, s) = \alpha(s, 0)\gamma_0\alpha(t, 0)' + \int_0^t \int_0^s \alpha(t, u)b(u)c_z(u, v)'b(v)'\alpha(s, v)'dudv$$

by expectation, so that

$$\begin{aligned}
\frac{\partial c(t, s)}{\partial t} &= a(t)c(t, s) + d(t, s), \\
\text{where } d(t, s) &= b(t) \int_0^s c_z(t, v)b(v)'\alpha(s, v)'dv,
\end{aligned} \tag{7.14}$$

by differentiation with respect to  $t$  (Exercise 7.6). The latter equation can be solved for the initial condition  $\gamma(s) = c(s, s)$  that can be determined from

$$\dot{\gamma}(t) = a(t)\gamma(t) + \gamma(t)a(t)' + d(t, t) + d(t, t)' \tag{7.15}$$

with the initial condition  $\gamma(0) = \gamma_0$ .

For the real-valued process  $Z(t) = Y(t)^2$  considered in this example, we find  $c_z(t, s) = E[Y(t)^2Y(s)^2] - E[Y(t)^2]E[Y(s)^2] = 2\exp(-2\alpha(t-s))$  by using properties of Gaussian variables. The covariance function in (7.12) is obtained by solving the differential equations  $\dot{\gamma}(t) = -2\rho\gamma(t) + 2d(t, t)$  and  $\partial c(t, s)/\partial t = -\rho c(t, s) + d(t, s)$ , where  $d(t, s) = 2(e^{-2\alpha(t-s)} - e^{-2\alpha t - \rho s})/(2\alpha + \rho)$ .

We now use direct calculations to find the expression of  $c(t, s)$ . The integral form of the equation defining  $X(t)$  is

$$X(t) - X(s) = -\rho \int_s^t X(u)du + \int_s^t Y(u)^k du, \quad t > s,$$

which gives  $r(t, s) - r(s, s) = -\rho \int_s^t r(u, s)du + \int_s^t E[Y(u)^k X(s)]du$  by multiplication with  $X(s)$  and taking the expectation, where  $r(t, s) = E[X(t)X(s)]$  and  $k \geq 1$  is an integer. This equation gives

$$\frac{\partial r(t, s)}{\partial t} = -\rho r(t, s) + E[Y(t)^k X(s)], \quad t > s,$$

by differentiation with respect to  $t$ . For  $t > s$ , we have  $Y(t) = e^{-\alpha t} \left( Y(s)e^{\alpha s} + \sqrt{2\alpha} I(s, t) \right)$  with  $I(s, t) = \int_s^t e^{\alpha u} dB(u)$ , so that

$$\begin{aligned} E[Y(t)^k X(s)] &= E \left[ e^{-\alpha k t} \left( Y(s)e^{-\alpha s} + \sqrt{2\alpha} I(s, t) \right)^k X(s) \right] \\ &= e^{-\alpha k t} \sum_{q=0}^k \frac{k!}{q!(k-q)!} e^{\alpha(k-q)s} (2\alpha)^{q/2} E[Y(s)^{k-q} X(s)] E[I(s, t)^q] \end{aligned}$$

since  $I(s, t)$  is independent of  $X(s)$  and  $Y(s)$ . The expectation  $E[Y(s)^{k-q} X(s)] = \mu(1, k-q; s)$  has been calculated in Example 7.3. The variable  $I(s, t)$  is Gaussian with mean 0 and variance  $\int_s^t e^{2\alpha u} du = (e^{2\alpha t} - e^{2\alpha s})/(2\alpha)$  by Itô's isometry (Sect. 4.4). We have  $E[Y(t)^2 X(s)] = e^{-2\alpha t} [e^{2\alpha s} \mu(1, 2; s) + 2\alpha \mu(1, 0; s) E[I(s, t)^2]]$ , so that  $E[Y(t)^2 X(s)] = (\mu(1, 2) - \mu(1, 0))e^{-2\alpha(t-s)} + \mu(1, 0)$  in the stationary regime, which yields (7.12) since  $c(t, s) = r(t, s) - E[X(t)]E[X(s)]$ ,  $\mu(1, 0) = 1/\rho$ , and  $\mu(1, 2) = (21\alpha + 3\rho)/[\rho(2\alpha + \rho)]$ . ▲

*Example 7.5* The characteristic function  $\varphi(u, v; t) = E[e^{i(uX(t) + vY(t))}]$  of  $(X(t), Y(t))$  in the previous example satisfies the partial differential equation

$$\frac{\partial \varphi}{\partial t} = -\rho u \frac{\partial \varphi}{\partial u} + (-1)^k t^{k+1} u \frac{\partial^k \varphi}{\partial v^k} - \alpha v \frac{\partial \varphi}{\partial v} - \alpha v^2 \varphi$$

with boundary conditions resulting from the equality of  $\partial^{p+q} \varphi(u, v; t) / \partial u^p \partial v^q$  at  $(u = 0, v = 0)$  and  $i^{p+q} E[X(t)^p Y(t)^q]$ . The moments of  $(X(t), Y(t))$  are from Example 7.3, and the initial condition results from properties of  $(X(0), Y(0))$ . ◇

*Proof* Itô's formula applied to  $\exp(i(uX(t) + vY(t)))$  with  $(u, v) \in \mathbb{R}^2$  gives the partial differential for  $\varphi(u, v; t)$  by averaging and differentiation with respect to time. Similar calculations have been performed in Example 5.15. ▲

### 7.2.3 Continuous Time Nonlinear Systems

Let  $X(t)$ ,  $t \geq 0$ , be an  $\mathbb{R}^d$ -valued stochastic process satisfying the stochastic differential and integral equations

$$\begin{aligned} dX(t) &= a(X(t-)) dt + b(X(t-)) dY(t) \quad \text{and} \\ X(t) &= X(0) + \int_0^t a(X(s-)) ds + \int_0^t b(X(s-)) dY(s), \quad t \geq 0, \end{aligned} \quad (7.16)$$

where  $a$  and  $b$  are  $(d, 1)$  and  $(d, d')$  matrices whose entries are real-valued Borel measurable functions and  $Y(t)$  denotes an  $\mathbb{R}^{d'}$ -valued semimartingale (Sect. 5.5.1). It is assumed that the drift and diffusion coefficients  $a$  and  $b$  are such that the solution of (7.16) exists and is unique (Theorems 5.7, 5.8, 5.11, and 5.12).

We develop equations for moments, distributions, and other properties of  $X(t)$  by Itô's formula for continuous and arbitrary semimartingale in Theorems 5.3 and 5.4. The differential equation for the density of  $X(t)$  is referred to as the Fokker–Planck equation.

Let  $\mu(q_1, \dots, q_d; t) = E\left[\prod_{k=1}^d X_k(t)^{q_k}\right]$  be moments of order  $q = q_1 + \dots + q_d$ , where  $q_k \geq 0$  are integers,  $\varphi(u; t) = E[\exp(iu'X(t))]$ ,  $u = (u_1, \dots, u_d) \in \mathbb{R}^d$ , be the marginal characteristic function, and  $f(x; t | x_0; 0) = \int_{\mathbb{R}^d} e^{-iu'x} \varphi(u; t) du / (2\pi)^d$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ , be the marginal density of  $X(t)$  at time  $t \geq 0$  for a specified initial state  $X(0) = x_0$  assumed to be independent of the driving noise.

**Theorem 7.3** *The moments and the characteristic functions of  $X(t)$  in (7.16) driven by an  $\mathbb{R}^{d'}$ -valued Brownian motion  $Y(t) = B(t)$  satisfy the differential equations*

$$\begin{aligned} \dot{\mu}(q_1, \dots, q_d; t) &= \sum_{k=1}^d E\left[a_k(X(t)) \frac{\partial g(X(t))}{\partial x_k}\right] \\ &\quad + \frac{1}{2} \sum_{k,l=1}^d E\left[(b(X(t))b(X(t)))'_{kl} \frac{\partial^2 g(X(t))}{\partial x_k \partial x_l}\right] \text{ and} \\ \frac{\partial \varphi(u; t)}{\partial t} &= i \sum_{k=1}^d u_k E[a_k(X(t)) e^{iu'X(t)}] - \frac{1}{2} \sum_{k,l=1}^d u_k u_l E\left[(b(X(t))b(X(t)))'_{kl} e^{iu'X(t)}\right] \end{aligned} \quad (7.17)$$

where  $g(X(t)) = \prod_{k=1}^d X_k(t)^{q_k}$ .

*Proof* The expectations of Itô's formula applied to  $g(X(t))$  and  $\exp(iu'X(t))$  give (7.17) by differentiation with respect to time. For example, Itô's formula in (5.15) applied to  $\exp(iu'X(t))$  gives

$$e^{iu'X(t)} - e^{iu'X(0)} = \sum_{k=1}^d i \int_0^t u_k e^{iu'X(s)} dX_k(s) - \frac{1}{2} \sum_{k,l=1}^d \int_0^t u_k u_l e^{iu'X(s)} d[X_k, X_l](s),$$

which yields the second formula in (7.17) by expectation and differentiation with respect to time since  $dX_k(t) = a_k(X(s))ds + \sum_{r=1}^{d'} b_{kr}(X(s))dB_r(s)$  and  $d[X_k, X_l](s) = \sum_{r=1}^{d'} b_{kr}(X(s))b_{lr}(X(s))ds = (b(X(s))b(X(s)))'_{kl}ds$ .

Note that (7.17) delivers differential equations for the moments and the characteristic function of  $X(t)$  only if the drift and diffusion coefficients in (7.16) are polynomials of  $X(t)$ . In this case, the expectations in (7.17) can be expressed as moments of  $X(t)$  and partial derivatives of  $\varphi(u; t)$ .  $\blacktriangle$

Suppose the drift and diffusion coefficients are polynomials of  $X(t)$ . The solution of the ordinary differential equation for  $\mu(q_1, \dots, q_d; t)$  only requires initial values  $\mu(q_1, \dots, q_d; 0)$  that can be calculated from the initial state  $X(0)$ , which is known. Generally, it is not possible to find the moments of  $X(t)$  exactly since the moment equations in (7.17) form an infinite hierarchy, as illustrated by a subsequent example. The solution of the partial differential equation for  $\varphi(u; t)$  requires initial and

boundary conditions. Initial conditions result from properties of  $X(0)$ . The boundary condition  $\varphi(0; t) = 1$  follows from the definition of the characteristic function. Additional boundary conditions needed for solution result from the following fact. If  $h : \mathbb{R} \rightarrow \mathbb{R}$  is integrable, then  $\zeta(u) = \int_{\mathbb{R}} e^{iu x} h(x) dx \rightarrow 0$  as  $|u| \rightarrow \infty$  ([25], Lemma 3, p. 513). This implies  $\varphi(u; t) \rightarrow 0$  as  $\|u\| \rightarrow \infty$  if  $X(t)$  has a density and  $\partial^q \varphi(u; t) / [\partial u_1^{q_1} \cdots \partial u_d^{q_d}] \rightarrow 0$  as  $\|u\| \rightarrow \infty$  if  $X(t)$  has finite moments of order  $q = q_1 + \cdots + q_d$ . Since  $\|u\| \rightarrow \infty$  holds if at least one of the coordinates of  $u$  converges to infinity, say  $u_d$ , and

$$\varphi(u; t) = \int_{\mathbb{R}^{d-1}} e^{i \sum_{j=1}^{d-1} u_j x_j} f(\hat{x}) \left[ \int_{\mathbb{R}} e^{iu_d x_d} f(x_d | \hat{x}) dx_d \right] d\hat{x},$$

we have  $\int_{\mathbb{R}} e^{iu_d x_d} f(x_d | \hat{x}) dx_d \rightarrow 0$  as  $|u_d| \rightarrow \infty$  since  $f(x_d | \hat{x})$  is integrable, where  $\hat{x} = (x_1, \dots, x_{d-1})$ ,  $f(\hat{x})$  denotes the density of  $\hat{X} = (X_1, \dots, X_{d-1})$ , and  $f(\cdot | \hat{x})$  is the density of  $X_d | (\hat{X} = \hat{x})$ .

*Example 7.6* Let  $X(t)$  be the solution of  $dX(t) = (\beta X(t) - X(t)^3)dt + \sigma dB(t)$ ,  $t \geq 0$ , with initial state  $X(0)$  independent of  $B(t)$ , where  $\beta, \sigma$  are real constants. The moments  $\mu(q; t) = E[X(t)^q]$  satisfy the ordinary differential equation

$$\dot{\mu}(q; t) = q\beta\mu(q; t) - q\mu(q+2; t) + \frac{q(q-1)\sigma^2}{2}\mu(q-2; t), \quad t \geq 0$$

with  $\mu(q; 0) = E[X(0)^q]$ . It is not possible to find the moments of  $X(t)$  exactly since the moment equations for  $X(t)$  form an infinite hierarchy. For example,  $\dot{\mu}(1; t) = \beta\mu(1; t) - \mu(3; t)$  depends on  $\mu(3; t)$ , which is not known. Closure methods have been proposed to solve approximately moment equations for nonlinear systems. The methods are heuristic, and can be unsatisfactory [26, 27].  $\diamond$

*Example 7.7* The stationary density of  $X(t)$  in the previous example is an even function, so that all stationary odd order moments are zero. The non-zero stationary moments  $\mu(q) = \lim_{t \rightarrow \infty} \mu(q; t)$  satisfy the recurrence formula

$$\mu(2(k+1)) = a\mu(2k) + (2k-1)b\mu(2(k-1)), \quad k = 1, 2, \dots,$$

where  $a = \beta$  and  $b = \sigma^2/2$ . This formula gives  $\mu(4) = a\mu(2) + b$ ,  $\mu(6) = (a^2 + 3b)\mu(2) + ab$ ,  $\mu(6) = (a^3 + 8ab)\mu(2) + a^2b + 5b^2$ , and so on. Closure methods retain a finite number of moment equations and supplement them with additional relations between the unknown moments in these equations, so that the augmented set of equations can be solved. For example, let  $k_0$  be an arbitrary closure level and  $\mu(2(k_0+1)) = \zeta\mu(2k_0)$  a closure technique, where  $\zeta > 0$  is a constant.

For  $\beta = -1$  and  $\sigma = 1$ ,  $\mu(2)$  takes values in  $I_2 = [0, 1/2]$  for  $k_0 = 2$  and  $I_4 = [0, 1/2]$  for  $k_0 = 4$ . It can be shown that  $\lim_{k_0 \rightarrow \infty} I_{k_0} = \{\mu(2)\}$  irrespective of the value of  $\zeta$  [27]. For  $\beta = 1$  and  $\sigma = 1$ ,  $\mu(2)$  takes values in  $(0, \infty)$  irrespective of the closure level and the value of  $\zeta$ . This shows that the particular closure technique used for solution is irrelevant. The structure of the stochastic differential equation determines the success or the failure of closure techniques [26, 27].  $\diamond$

**Example 7.8** The characteristic function  $\varphi(u; t) = E[\exp(iuX(t))]$  of  $X(t)$  in Example 7.6 satisfies the partial differential equation

$$\frac{\partial \varphi(u; t)}{\partial t} = \beta u \frac{\partial \varphi(u; t)}{\partial u} + u \frac{\partial^3 \varphi(u; t)}{\partial u^3} + \frac{\sigma^2 u^2}{2} \varphi(u; t), \quad u \in \mathbb{R},$$

with the initial condition  $\varphi(u; 0) = E[\exp(iuX(0))]$ . The boundary conditions can be  $\varphi(0; t) = 1$ ,  $\lim_{|u| \rightarrow \infty} \varphi(u; t) = 0$ , and  $\lim_{|u| \rightarrow \infty} \partial \varphi(u; t) / \partial u = 0$ . Alternative boundary conditions are available in this case since  $X(t) \stackrel{d}{=} -X(t)$  so that the density of  $X(t)$  is an even function implying  $E[X(t)] = 0$  provided it exists,  $\varphi(u; t) = \varphi(-u; t)$ , and  $\varphi(u; t) \in \mathbb{R}$ . Hence, it is sufficient to solve the partial differential equation for  $\varphi(u; t)$  in  $[0, \infty)$  with the boundary conditions  $\varphi(0; t) = 1$ ,  $\partial \varphi(u; t) / \partial u = 0$  at  $u = 0$ , and  $\lim_{|u| \rightarrow \infty} \varphi(u; t) = 0$ . The latter conditions can be replaced for numerical calculations with  $\varphi(a; t) = 0$  for a sufficiently large  $a > 0$ .  $\diamond$

In summary, differential equations can be obtained for the moments and characteristic function of  $X(t)$  defined by (7.16) provided the entries of the drift and diffusion coefficients are polynomials of  $X(t)$ . Generally, the moment equations cannot be solved exactly since they form an infinite hierarchy, a very different situation from that of moment equations for the state of linear systems driven by polynomials of filtered Gaussian and/or Poisson processes (Example 7.3). The solution of the partial differential equation for the characteristic function requires both initial and boundary conditions. The specification of boundary conditions beyond  $\varphi(0; t) = 1$  may pose difficulties since they involve properties of  $X(t)$ , that are not known. Numerical solutions are usually obtained under the assumption that the support of  $\varphi(u; t)$  in the argument  $u$  is a bounded rectangle  $R$  in  $\mathbb{R}^d$ . Boundary conditions on  $\varphi(u; t)$  and its derivatives for  $\|u\| \rightarrow \infty$  are usually imposed on the boundary of  $R$ . For the process in Example 7.8 this means to replace  $\lim_{|u| \rightarrow \infty} \varphi(u; t) = 0$  with  $\varphi(\bar{u}; t) = 0$ ,  $\bar{u} > 0$ . Solutions for increasing value of  $\bar{u}$  need to be performed to determine whether the assumption that  $X(t)$  has finite moments is valid and/or  $R = (-\bar{u}, \bar{u})$  is sufficiently large. Stable solutions for increasing values of  $\bar{u}$  would indicate that the moments of  $X(t)$  considered in the analysis are finite and that  $R = (-\bar{u}, \bar{u})$  has an adequate size.

The following theorem shows that the Fourier transform of the second differential equation in (7.17) is a partial differential equation for the density  $f(x; t | x_0; 0)$  of  $X(t) | (X(0) = x_0)$ , referred to as the Fokker–Planck equation. The entries of the drift and diffusion coefficients are not required to be polynomials of  $X(t)$ .

**Theorem 7.4** *If the conditions  $a_k(x)f(x; t | x_0; 0) \rightarrow 0$ ,  $(b(x)b(x'))_{kl}f(x; t | x_0; 0) \rightarrow 0$ , and  $\partial[(b(x)b(x'))_{kl}f(x; t | x_0; 0)]/\partial x_k \rightarrow 0$ ,  $k, l = 1, \dots, d$ , as  $\|x\| \rightarrow \infty$  hold, and if  $f(x; t | x_0; 0)$  and  $\partial f(x; t | x_0; 0)/\partial t$  are continuous, then  $f(x; t | x_0; 0)$  satisfies the partial differential equation*

$$\frac{\partial f}{\partial t} = - \sum_{k=1}^d \frac{\partial}{\partial x_k} [a_k(x)f] + \frac{1}{2} \sum_{k,l=1}^d \frac{\partial^2}{\partial x_k \partial x_l} [(b(x)b(x'))_{kl}f] \quad (7.18)$$

with the initial condition  $f(x; 0 \mid x_0; 0) = \delta(x - x_0)$  and boundary conditions depending on the objective of the analysis, that are discussed later in this section.

*Proof* The first three conditions impose constraints on the tails of  $f(x; t \mid x_0; 0)$ , and relate to the behavior of the characteristic function  $\varphi(u; t)$  of  $X(t)$  around  $u = 0$ . We show that the Fourier transform of (7.18) coincides with the second formula in (7.17).

The integral of the left side of (7.18) multiplied by  $\exp(iu'x)$  is

$$\int_{\mathbb{R}^d} e^{iu'x} \frac{\partial f(x; t \mid x_0; 0)}{\partial t} dx = \frac{\partial}{\partial t} \int_{\mathbb{R}^d} e^{iu'x} f(x; t \mid x_0; 0) dx = \frac{\partial \varphi(u; t)}{\partial t},$$

by using Leibnitz's rule in  $R \times [0, \tau]$ , that applies since  $e^{iu'x} \partial f / \partial t$  and  $e^{iu'x} f$  are continuous in  $\mathbb{R}^d \times [0, \tau]$ , where  $R$  is an arbitrary but bounded rectangle in  $\mathbb{R}^d$ ,  $[0, \tau]$  denotes a bounded range for  $t$ , and  $f$  is a short hand notation for  $f(x; t \mid x_0; 0)$ .

The Fourier transform of the first term on the right side of (7.18) is

$$\begin{aligned} - \int_{\mathbb{R}^d} e^{iu'x} \frac{\partial(a_k f)}{\partial x_k} dx &= - \int_{\mathbb{R}^{d-1}} \prod_{r=1, r \neq k}^d e^{iu_r x_r} dx_r \int_{\mathbb{R}} e^{iu_k x_k} \frac{\partial(a_k f)}{\partial x_k} dx_k \\ &= iu_k \int_{\mathbb{R}^d} e^{iu'x} a_k f dx = iu_k E[e^{iu'X(t)} a_k(X(t))] \end{aligned}$$

since  $\int_{\mathbb{R}} e^{iu_k x_k} \partial(a_k f) / \partial x_k dx_k = e^{iu_k x_k} a_k f \big|_{-\infty}^{\infty} - \int_{\mathbb{R}} iu_k e^{iu'x} a_k f dx_k$  by integration by parts and  $e^{iu_k x_k} a_k f \big|_{-\infty}^{\infty} = 0$  by assumption. Similar calculations show that the integral of the second term on the right side of (7.18) multiplied by  $\exp(iu'x)$  over  $\mathbb{R}^d$  coincides with the second term on the right side of (7.17).  $\blacktriangle$

An alternative form of (7.18) is

$$\frac{\partial f}{\partial t} = - \sum_{k=1}^d \frac{\partial \lambda_k(x; t)}{\partial x_k}, \quad \text{where } \lambda_k(x; t) = a_k f - \frac{1}{2} \sum_{l=1}^d \frac{\partial((bb')_{kl} f)}{\partial x_l}. \quad (7.19)$$

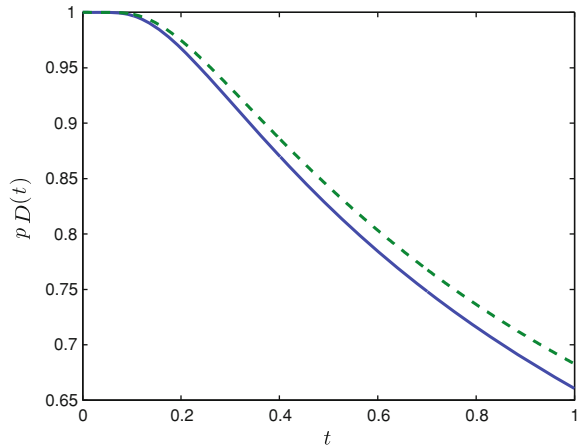
The vector  $\lambda(x; t) \in \mathbb{R}^d$  with coordinates  $\{\lambda_k(x; t)\}$  is referred to as probability current. Let  $p_D(t) = \int_D f(x; t \mid x_0; 0) dx$ , where  $D$  is on open subset of  $\mathbb{R}^d$  with boundary  $\partial D$ . Then

$$\frac{\partial p_D(t)}{\partial t} = - \sum_{k=1}^d \int_D \frac{\partial \lambda_k(x; t)}{\partial x_k} dx = - \int_{\partial D} \lambda(x; t) \cdot n(x) d\sigma(x) \quad (7.20)$$

by the divergence theorem ([28], p. 116), where  $n(x)$  denotes the exterior normal at  $x \in \partial D$  and  $d\sigma(x)$  is an infinitesimal surface element on  $\partial D$ . Note that  $-\lambda(x; t) \cdot n(x) d\sigma(x)$  represents the rate of change of  $p_D(t)$  caused by probability flow through the element of area  $d\sigma(x)$  of  $\partial D$ .

**Definition 7.2** Let  $T = \inf\{t \geq 0, X(t) \notin D, X(0) = x_0 \in D\}$  be a stopping time denoting the first time  $X(t)$  starting at  $x_0 \in D$  exists  $D$ . If  $T = \infty$  a.s.,  $\partial D$  is an

**Fig. 7.2** Analytical (dotted line) and finite difference (solid line) solutions for  $p_D(t)$



inaccessible boundary, that is called natural if  $X(t)$  never reaches it and attracting if  $\lim_{t \rightarrow \infty} X(t) \in \partial D$ . If  $T < \infty$  a.s.,  $\partial D$  is an accessible boundary, that is called reflecting if  $\lambda(x; t) \cdot n(x) = 0$ ,  $x \in \partial D$  and absorbing if  $f(x; t | x_0; 0) = 0$ ,  $x \in \partial D$ .

**Example 7.9** The solution  $X(t)$  of the stochastic differential equation  $dX(t) = cX(t)dt + \sigma X(t)dB(t)$ ,  $t \geq 0$ , referred to as the geometric Brownian motion, is  $X(t) = x_0 \exp[(c - \sigma^2/2)t + \sigma dB(t)]$ , where  $X(0) = x_0$  denotes the initial state (Example 5.11). We have  $\lim_{t \rightarrow \infty} X(t) = X(0) \exp[(c - \sigma^2/2)t]$  since  $B(t)/t \rightarrow 0$  a.s., as  $t \rightarrow \infty$  ([29], Sect. 6.4). If  $X(0) \neq 0$  and  $c - \sigma^2/2 < 0$ , then  $x = 0$  is an inaccessible, attracting boundary.  $\diamond$

**Example 7.10** The density  $f(x; t)$  of a Brownian motion  $B(t)$ ,  $t \geq 0$ , starting at  $B(0) = 0$  satisfies the Fokker–Planck equation  $\partial f(x; t)/\partial t = (1/2)\partial^2 f(x; t)/\partial x^2$  with the initial condition  $f(x; 0) = \delta(x)$ . Let  $T = \inf\{t \geq 0 : B(t) \geq a\}$ ,  $a > 0$ , be the first time  $B(t)$  exists  $D = (-\infty, a)$ . The probability that  $B(t)$  does not exist  $D$  in a time interval  $[0, t]$  is  $p_D(t) = P(T > t) = 2\Phi(a/\sqrt{t}) - 1$ . This probability can also be obtained by solving the Fokker–Planck equation for  $f(x; t)$  numerically with the absorbing boundary  $f(a; t) = 0$ . Figure 7.2 shows with dotted and solid lines the analytical expression of  $p_D(t)$  and a finite difference solution of  $\partial f(x; t)/\partial t = (1/2)\partial^2 f(x; t)/\partial x^2$  under appropriate boundary conditions.  $\diamond$

**Proof** We have  $P(B(t) \geq a) = P(B(t) \geq a | T \leq t)P(T \leq t) + P(B(t) \geq a | T > t)P(T > t) = P(B(t) \geq a | T \leq t)P(T \leq t)$  since  $P(B(t) \geq a | T > t) = 0$ . If  $T \leq t$ , the Brownian motion has reached  $x = a$  at a time prior to  $t$ . The symmetry of the Brownian motion implies that the events  $\{B(t) \geq a | T \leq t\}$  and  $\{B(t) \leq a | T \leq t\}$  are equally likely, so that  $P(B(t) \geq a) = P(T \leq t)/2$  or  $P(T \leq t) = 2P(B(t) \geq a) = 2\Phi(-a/\sqrt{t})$ .  $\blacktriangle$

*Example 7.11* The partial differential equations for the characteristic and density functions of  $X(t)$  defined by  $dX(t) = -\rho X(t)dt + dB(t)$  and  $dX(t) = -\rho X(t)dt + dC(t)$  are

Gaussian noise:

$$\frac{\partial \varphi}{\partial t} = -\rho u \frac{\partial \varphi}{\partial u} - \frac{u^2}{2} \varphi \quad \frac{\partial f}{\partial t} = \rho \frac{\partial(xf)}{\partial x} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}$$

Poisson noise:

$$\frac{\partial \varphi}{\partial t} = -\rho u \frac{\partial \varphi}{\partial u} + \lambda (\varphi_{Y_1}(u) - 1) \varphi \quad \frac{\partial f}{\partial t} = \rho \frac{\partial(xf)}{\partial x} + \lambda \sum_{r=1}^{\infty} \frac{(-1)^r E[Y_1^r]}{r!} \frac{\partial^r f}{\partial x^r},$$

where  $\rho > 0$  is a constant,  $B(t)$  denotes a Brownian motion,  $C(t) = \sum_{k=1}^{N(t)} Y_k$ ,  $N(t)$  is a Poisson process with intensity  $\lambda > 0$ , and  $\{Y_k\}$  are iid random variables with bounded moments of any order, and  $\varphi_{Y_1}$  is the characteristic function of  $Y_1$  (Exercise 7.9).  $\diamond$

*Example 7.12* Suppose the real-valued diffusion process  $X(t)$  defined by  $dX(t) = a(X(t))dt + b(X(t))dB(t)$  admits a stationary solution. The stationary density of  $X(t)$  is  $f(s) \propto \exp[2\beta(x)]/b(x)^2$ , where  $\beta'(x) = a(x)/b(x)^2$ . Additional examples can be found in [30] (Chap. 5) and [4] (Sect. 7.3.1).  $\diamond$

*Proof* The stationary version of (7.18) is  $d[a(x)f(x) + d(b(x)^2 f(x))/dx]/dx = 0$  so that  $a(x)f(x) + d(b(x)^2 f(x))/dx$  is a constant that must be zero under the assumptions in Theorem 7.4. The solution of  $a(x)f(x) + d(b(x)^2 f(x))/dx = 0$  is the stationary density of  $X(t)$ . For example,  $f(x) \propto \exp[(\alpha x^2 + \beta x^4)/2]/\sigma^2$ ,  $x \in \mathbb{R}$ , for  $a(x) = \alpha x + \beta x^3$  and  $b(x) = \sigma$  is a constant. There is no stationary density for  $\alpha > 0$  and  $\beta > 0$ . If  $\alpha > 0$  and  $\beta < 0$ , the stationary density  $f(x)$  exists and has two modes.  $\blacktriangle$

The dimension of  $X(t)$  is large in most applications, so that it is difficult to find its properties by solving the differential equations in (7.17) and (7.18). A broad range of approximate methods has been proposed to characterize  $X(t)$ , for example, perturbation, Taylor series, Neumann series, equivalent linearization, stochastic averaging, and other methods ([30], Chap. 6, [4], Sect. 4.9.4, [31]).

### 7.3 Stochastic Difference Equations with Random Coefficients

Difference equations can be obtained from stochastic differential equations with random coefficients by time discretization or can be constructed directly as discrete time models for various phenomena ([32], Sect. 1.1). Following some general considerations (Sect. 7.3.1), we discuss methods for solving stochastic difference equations with arbitrary random coefficients (Sects. 7.3.2–7.3.5) and random coefficients with small uncertainty (Sects. 7.3.6–7.3.7).



### 7.3.1 General Considerations

Let

$$X_{n+1} = X_n + a(X_n, Y_n, n\Delta t)\Delta t + b(X_n, Y_n, n\Delta t)W_n, \quad n = 0, 1, \dots, \quad (7.21)$$

be a discrete time version of (7.1) obtained by approximating the differentials in this equation by forward finite differences, where  $\Delta t > 0$  denotes the time step,  $X_n = X(n\Delta t)$ ,  $Y_n = Y(n\Delta t)$ ,  $S(t) = B(t)$ , and  $W_n = B((n+1)\Delta t) - B(n\Delta t)$  is an  $d'$ -dimensional vector with independent  $N(0, \Delta t)$  coordinates. It is assumed that the drift and the diffusion coefficients in (7.1) are nearly constant in  $\Delta t$ . The time series  $\{X_n\}$  conditional on  $\{Y_0, Y_1, \dots\}$  is a discrete time, continuous state Markov process whose transition probabilities can be obtained from the observation that  $X_{n+1} | X_n$  is a Gaussian vector with mean  $X_n + a(X_n, Y_n, t)\Delta t$  and covariance matrix  $b(X_n, Y_n, n\Delta t)E[W_n W_n']b(X_n, Y_n, n\Delta t)'$ .

The recurrence formula in (7.21) written in the form

$$X_{n+1} = A_n(X_n) + B_n(X_n)W_n, \quad n = 0, 1, 2, \dots, \quad (7.22)$$

defines an autoregressive model with random coefficients, where  $A_n(X_n) = X_n + a(X_n, Y_n, n\Delta t)\Delta t$  and  $B_n(X_n) = b(X_n, Y_n, n\Delta t)$ . The model in (7.22) is said to be linear with multiplicative noise if  $A_n(X_n)$  and  $B_n(X_n)$  are linear in  $X_n$ , and linear with additive noise if  $A_n(X_n)$  is linear in  $X_n$  and  $B_n(X_n)$  does not depend on the state.

*Example 7.13* Let  $X(t)$  be a real-valued process satisfying the linear equation

$$dX(t) = a(Y(t), t)X(t)dt + b(Y(t), t)X(t)dB(t), \quad t \geq 0. \quad (7.23)$$

If  $a(Y(t), t)$  and  $b(Y(t), t)$  are replaced by deterministic constants, then  $X(t)$  is a geometric Brownian motion. If  $a(Y(t), t)$  and  $b(Y(t), t)X(t)$  are replaced by deterministic constants, then  $X(t)$  is an Ornstein–Uhlenbeck process. The discrete time version of (7.23) is

$$X_{n+1} = A_n X_n + B_n X_n W_n, \quad n = 0, 1, \dots, \quad (7.24)$$

where  $A_n$  and  $B_n$  are random matrices depending on  $Y(t)$ .  $\diamond$

Most studies on difference equations with random coefficients are limited to linear models with additive noise and state independent coefficients. Autoregressive models with random coefficients and state  $X_n \in \mathbb{R}$  defined by

$$X_{n+1} = A_n X_n + W_n, \quad n = 0, 1, \dots, \quad (7.25)$$

have been studied extensively, where  $\{A_n\}$  denotes a real-valued random sequence that is independent of  $\{W_n\}$  and  $X_0$ , and  $\{W_n\}$  is an iid sequence. For example, conditions for the existence of the stationary solution and other asymptotic properties

of  $\{X_n\}$  in (7.25) are established in [33]. Moment equations can be found in [12] for the vector version of (7.25) without driving noise under the assumption that  $\{A_n\}$  is a Markov chain with a finite number of states. In this case, the augmented state  $(X_n, A_n)$  is Markov chain whose evolution in time is defined by a nonlinear recurrence formula driven by Gaussian white noise, so that it is not Gaussian. Properties of the state of first order autoregressive models with random coefficients whose distributions depend on uncertain parameters are examined in [13, 34, 35] within a Bayesian framework. The reminder of this section summarizes some of the results in these references and presents some related facts.

**Theorem 7.5** *If  $\{A_n\}$  in (7.25) are iid random variables with  $|A_0| < 1$  a.s. that are independent of the iid sequence  $\{W_n\}$  with mean 0 and variance 1, and  $\{A_n\}$  and  $\{W_n\}$  are independent of  $X_0$ , then*

$$\begin{aligned}\mu_n &= E[X_n] = (E[A_0])^n \mu_0 \rightarrow 0, n \rightarrow \infty \\ \gamma_n &= \text{Var}[X_n] = (E[A_0^2])^n \gamma_0 + \sum_{i=0}^{n-1} (E[A_0^2])^i \\ &\quad + \mu_0^2 \text{Var}[A_0] \sum_{i=0}^{n-1} (E[A_0])^{2(n-i+1)} (E[A_0^2])^i \rightarrow \frac{1}{1 - E[A_0^2]}, n \rightarrow \infty \\ c(p, q) &= \text{Cov}[X_p, X_q] = (E[A_0])^{|p-q|} \gamma_{p \wedge q} \rightarrow \frac{(E[A_0])^{|p-q|}}{1 - E[A_0^2]}, p, q \rightarrow \infty,\end{aligned}\tag{7.26}$$

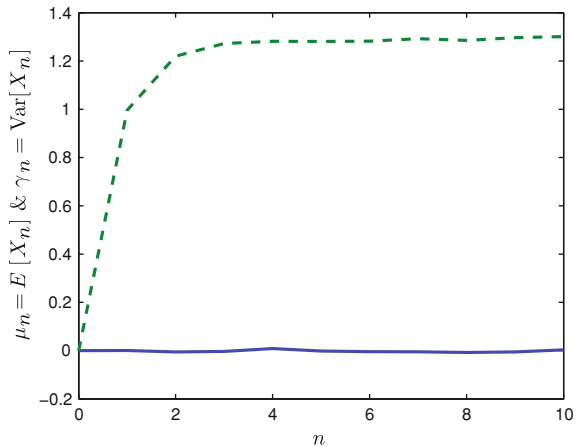
so that  $\{X_n\}$  becomes weakly stationary as time increases indefinitely.

*Proof* The expectation of (7.25) gives  $\mu_{n+1} = E[X_{n+1}] = E[A_n X_n] = E[A_n] E[X_n] = E[A_0] \mu_n$  since  $X_n$  is a function of  $(A_{n-1}, \dots, A_0, W_{n-1}, \dots, W_0)$  and  $E[W_n] = 0$ . Repeated applications of the recurrence formula  $\mu_{n+1} = E[A_0] \mu_n$  gives the expression of  $\mu_n$ . We have  $\lim_{n \rightarrow \infty} \mu_n = 0$  since  $|A_0| < 1$  a.s., so that  $E[A_0] < 1$ .

Subtract  $\mu_{n+1} = E[A_0] \mu_n$  from (7.25) to obtain  $\tilde{X}_{n+1} = A_n \tilde{X}_n + \tilde{A}_n \mu_n + W_n$ , where  $\tilde{X}_n = X_n - E[X_n]$  and  $\tilde{A}_n = A_n - E[A_n]$ . The expectation of the square of this equation gives the recurrence formula  $\gamma_{n+1} = E[A_0^2] \gamma_n + \text{Var}[A_0] \mu_n^2 + 1$  since  $E[W_n^2] = 1$ ,  $E[A_n \tilde{A}_n \tilde{X}_n] = 0$ ,  $E[A_n \tilde{X}_n W_n] = 0$ , and  $E[\tilde{A}_n W_n] = 0$  by assumptions. Repeated applications of the recurrence formula for  $\gamma_n$  give the expression of the variance of  $X_n$  in (7.26). The first term in the expression of  $\gamma_n$  converges to 0 as  $n \rightarrow \infty$  since  $E[A_0^2] < 1$ . The second term is a geometric progression with sum  $1/(1 - E[A_0^2])$  as  $n \rightarrow \infty$ . Let  $S_{n-1}$  denote the summation in the third term. Since  $0 \leq S_{n-1} \leq \sum_{i=0}^{n-1} (E[A_0])^i$  and  $\sum_{i=0}^{n-1} (E[A_0])^i$  is convergent,  $\lim_{n \rightarrow \infty} S_n = S$  exists and is finite. We also have  $S_n = (E[A_0^2])^n + (E[A_0])^2 S_{n-1}$  implying  $S = 0 + (E[A_0])^2 S$  by taking the limit  $n \rightarrow \infty$ , so that  $S = 0$  since  $E[A_0] \neq 0$ .

Multiply  $\tilde{X}_{n+1} = A_n \tilde{X}_n + \tilde{A}_n \mu_n + W_n$  by  $\tilde{X}_n$ . The expectation of the resulting expression is  $c(n, n+1) = E[\tilde{X}_{n+1} \tilde{X}_n] = E[A_0] \gamma_n$ , that is, a formula for the lag 1

**Fig. 7.3** Estimates of mean and variance functions of  $X_n$  in solid and dotted line for  $(\rho', \rho'') = (-0.2, 0.9)$



covariance function at times  $n$  and  $n + 1$ . In a similar manner we find

$$c(p, q) = E[\tilde{X}_p \tilde{X}_q] = (E[A_0])^{|p-q|} \gamma_{p \wedge q},$$

which converges to the stated result as  $p, q \rightarrow \infty$ .  $\blacktriangle$

**Example 7.14** Suppose  $A_n$  in (7.25) is uniformly distributed in  $(\rho', \rho'')$ ,  $-1 < \rho' < \rho'' < 1$ , and  $W_n \sim N(0, 1)$ . Figure 7.3 shows with solid and dotted lines estimates of the mean  $\mu_n$  and the variance  $\gamma_n$  as a function of  $n$  for  $(\rho', \rho'') = (-0.2, 0.9)$  based on 100000 independent samples of  $X_n$ . The asymptotic values of the mean and variance in (7.26) are  $\lim_{n \rightarrow \infty} \mu_n = 0$  and  $\lim_{n \rightarrow \infty} \gamma_n = 1/(1 - E[A_0^2]) = 1.2875$ , in agreement with Monte Carlo estimates. In this illustration,  $\mu_n$  and  $\gamma_n$  converge rapidly to their stationary values.  $\diamond$

*Proof* The second moment of  $A_0$  is  $E[A_0^2] = \int_{\rho'}^{\rho''} u^2 du / (\rho'' - \rho') = ((\rho')^2 + \rho' \rho'' + (\rho'')^3)/3 = 0.2233$  for  $\rho' = -0.2$  and  $\rho'' = 0.9$ , so that  $\lim_{n \rightarrow \infty} \gamma_n = 1.2875$ . If  $A_n$  is deterministic, for example,  $\rho' = \rho'' = 0.9$ , the asymptotic mean and variance of  $X_n$  are  $\lim_{n \rightarrow \infty} \mu_n = 0$  and  $\lim_{n \rightarrow \infty} \gamma_n = 1/(1 - 0.9^2) = 5.26$ . This observation suggest that the variance of  $X_n$  is sensitive to the uncertainty in  $A_n$ .  $\blacktriangle$

**Theorem 7.6** Let  $\{X_n, n = 0, 1, \dots\}$  be an  $\mathbb{R}^d$ -valued random sequence defined by  $X_{n+1} = A_n X_n$ , where  $A_n$  are iid  $(d, d)$  random matrices. If the sequence  $\{A_n\}$  is independent of initial state  $X_0$ , the second moment properties of this sequence can be calculated from

$$\begin{aligned} E[X_{n+1}] &= E[A_n]E[X_n], \\ E[X_{n+1}X'_{n+1}] &= E\{A_n E[X_n X'_n] A'_n\}, \quad \text{and} \\ E[X_n X'_{n+p}] &= E[X_n X'_n] E[A'_n \cdots A'_{n+p-1}] \end{aligned} \quad (7.27)$$

for  $n = 0, 1, \dots$  and arbitrary integer  $p \geq 1$ .

*Proof* The expectation of  $X_{n+1} = A_n X_n$  is  $E[X_{n+1}] = E[A_n X_n] = E[A_n]E[X_n]$  since  $X_n$  is a function of  $(A_{n-1}, \dots, A_0, X_0)$  and  $A_n$  is independent of  $(A_{n-1}, \dots, A_0, X_0)$  by assumption. Similar considerations and properties of conditional expectation give  $E[X_{n+1}X'_{n+1}] = E\{E[A_n X_n(A_n X_n)' | A_n]\} = E\{A_n E[X_n X'_n] A'_n\}$ . Since  $X_{n+p} = A_{n+p-1} \cdots A_n X_n$ , we have  $E[X_n X'_{n+p}] = E[X_n(A_{n+p-1} \cdots A_n X_n)']$ , which gives the last formula in (7.27) by properties of  $X_n$ . The formulas in (7.27) have been derived in [12] following a different approach.  $\blacktriangle$

*Example 7.15* Suppose  $A_n$  in Theorem 7.6 is a stationary Markov chain with states the  $(d, d)$  deterministic matrices  $\{a_1, \dots, a_m\}$ , transition probabilities  $p_{sk} = P(A_n = a_k | A_{n-1} = a_s)$ , and stationary probability  $\pi_k = P(A_n = a_k)$ ,  $k = 1, \dots, m$ . The state expectation is  $E[X_{n+1}] = \sum_{k=1}^m a_k \pi_k \sum_{s=1}^m E[X_n | A_{n-1} = a_s] p_{ks}$ .  $\diamond$

*Proof* The probability of the event  $\{A_n = a_k, A_{n-1} = a_s\}$  is  $p_{sk}\pi_s$ . The expectation of  $X_{n+1} = A_n X_n$  conditional on this event is  $E[X_{n+1} | A_n = a_k, A_{n-1} = a_s] = a_k E[X_n | A_n = a_k, A_{n-1} = a_s]$  so that

$$\begin{aligned} E[X_{n+1}] &= \sum_{k,s=1}^m a_k E[X_n | A_n = a_k, A_{n-1} = a_s] p_{sk} \pi_s \\ &= \sum_{k,s=1}^m a_k E[X_n | A_{n-1} = a_s] p_{ks} \pi_k = \sum_{k=1}^m a_k \pi_k \sum_{s=1}^m E[X_n | A_{n-1} = a_s] p_{ks}, \end{aligned}$$

where we used  $p_{sk}\pi_s = p_{ks}\pi_k$  and the fact that  $X_n$  does not depend on  $A_n$ .  $\blacktriangle$

The random coefficients in (7.21) may model, for example, the degradation of a system's properties in time under exposure to random actions or simply our limited knowledge/information on the system behavior. In the later case, both the functional form of (7.21) and the properties of the random coefficients in this equation are uncertain. Prior information on both the functional form of (7.21) and features of its random coefficients can be incorporated within a Bayesian framework to identify an optimal functional form for (7.21) and corresponding distributions of the random parameters of this form. Let  $\{X_n\}$  be a real-valued autoregressive Gaussian sequence of unknown order  $k \geq 1$ , and consider a collection  $\mathcal{M}_k$ ,  $k = 1, 2, \dots, m$ , of such sequences whose members are defined by

$$X_{n+1} = \beta_0 + \beta_1 X_n + \cdots + \beta_k X_{n-k+1} + W_n, \quad (7.28)$$

where  $\tilde{\beta}_k = (\beta_0, \beta_1, \dots, \beta_k)'$  are unknown coefficients,  $k \geq 1$  is an integer, and  $\{W_n\}$  are independent  $N(0, 1/h_k)$  with unknown variance  $h_k > 0$ . Our objectives are to identify the optimal model for  $\{X_n\}$  and find distributions of its uncertain parameters based on (1) observations consisting of  $q > k$  consecutive, error free readings  $(z_1, \dots, z_q)$  of  $\{X_n\}$  and (2) prior information consisting of prior densities  $f'_k(\tilde{\beta}_k, h_k)$  of the uncertain parameters  $\tilde{\beta}_k$  and  $h_k$  in (7.28) and the prior probabilities  $p'_k \geq 0$  of the models  $\mathcal{M}_k$ ,  $k = 1, \dots, m$ , satisfying the conditions  $p'_k \geq 0$ ,  $k = 1, \dots, m$ , and  $\sum_{k=1}^m p'_k = 1$ . The densities of  $(\tilde{\beta}_k, h_k)$  that account for both prior information

and observations are denoted by  $f_k''(\tilde{\beta}_k, h_k)$  and are called posterior densities. Corresponding model probabilities are denoted by  $p_k''$  [36, 37]. An extensive discussion on the selection of prior densities can be found in [38] (Chaps. II and III). Physics and other knowledge can be used to select a set  $\mathcal{M}_k$ ,  $k = 1, 2, \dots, m$  of competing models and their prior probabilities.

The regression form of  $\mathcal{M}_k$  for the observation vector  $(z_1, \dots, z_q)$  is

$$V_k = \tilde{a}_k \tilde{\beta}_k + \frac{1}{\sqrt{h_k}} G_k, \quad (7.29)$$

where

$$\tilde{a}_k = \begin{bmatrix} 1 & z_k & \dots & z_1 \\ 1 & z_{k+1} & \dots & z_2 \\ \vdots & \vdots & & \vdots \\ 1 & z_{n-1} & \dots & z_{n-k} \end{bmatrix}, \quad (7.30)$$

$V_k = (X_{k+1}, \dots, X_n)'$ , and  $G_k$  is an  $(n - k)$ -dimensional vector with independent  $N(0, 1)$  entries.

**Theorem 7.7** *If  $f_k'$  is a normal-gamma density with parameters  $(\beta_k', \gamma_k', \rho_k', v_k')$  and  $n > k$ , then the posterior density  $f_k''$  of  $(\beta_k, h_k)$  is also a normal-gamma density with parameters  $(\beta_k'', \gamma_k'', \rho_k'', v_k'')$  given by*

$$\begin{aligned} (\gamma_k'')^{-1} &= (\gamma_k')^{-1} + a_k^T a_k \\ \tilde{\beta}_k'' &= \gamma_k'' \left[ (\gamma_k')^{-1} \tilde{\beta}_k' + a_k^T \tilde{z}_k \right] \\ \rho_k'' v_k'' &= \rho_k' v_k' + (\tilde{\beta}_k')^T (\gamma_k')^{-1} \tilde{\beta}_k' - (\tilde{\beta}_k')^T (\gamma_k'')^{-1} \tilde{\beta}_k'' + \tilde{z}_k^T \tilde{z}_k \\ v_k'' &= v_k' + n - k, \end{aligned} \quad (7.31)$$

where  $\tilde{z}_k = (z_{k+1}, \dots, z_n)$  is a column vector.

*Proof* Under our assumption on  $f_k'(\tilde{\beta}_k, h_k)$ , we have  $\tilde{\beta}_k | h_k \sim N(\tilde{\beta}_k', \gamma_k' / h_k)$  is a Gaussian vector with mean  $\tilde{\beta}_k'$  and covariance matrix  $\gamma_k' / h_k$  and  $h_k \sim G2(\rho_k', v_k')$  is gamma-2 distributed with parameters  $(\rho_k', v_k')$ , so that ([39], p. 226)

$$f_{h_k}'(\xi) = \frac{(\rho_k' v_k')^{v_k'/2}}{\Gamma(v_k'/2)} \xi^{v_k'/2-1} e^{-\rho_k' v_k' \xi / 2}.$$

Direct calculations using the Bayes formula give the result in (7.31) ([4], Sect. 9.8.2, [33], Sect. 3.2.3). The density  $f_k'$  is said to be a conjugate prior since it has the same functional form as  $f_k''$ . ▲

Arguments as in the above theorem can be used to find the posterior model probabilities  $\{p_k''\}$ . These probabilities can be used to identify optimal models [36, 37]. The random coefficients of an optimal model can be characterized by posterior densities

of the type in (7.31). The solution of the resulting difference equation with random coefficients can be obtained by the methods outlined in the following two sections.

Some of the methods for solving difference equations with deterministic coefficients and random input can be extended to solve a class of difference equations with random coefficients and input. Alternative methods have also been proposed to solve approximately this class of difference equations. The methods in Sects. 7.3.2–7.3.5 on Monte Carlo, conditional analysis, stochastic reduced order models, stochastic Galerkin, and stochastic collocation apply to difference equations with coefficients of arbitrary uncertainty. The Taylor and perturbation series methods in Sects. 7.3.6–7.3.7 are difference equations with coefficients that have small uncertainty.

### 7.3.2 Monte Carlo Simulation

As previously stated, Monte Carlo simulation is the only method that can be used to solve stochastic equations regardless of their size, structure, and complexity. The method involves the following three steps. First, independent samples of the random coefficients and driving noise need to be generated. Second, corresponding samples of the state vector need to be calculated. Third, resulting state samples are used to calculate state statistics.

*Example 7.16* Consider a mathematical pendulum with variable random length  $L(t)$ , and let  $X(t)$  be the angle measured anticlockwise between the vertical line and the thread holding the pendulum mass. Then  $X(t)$  is the solution of the differential equation

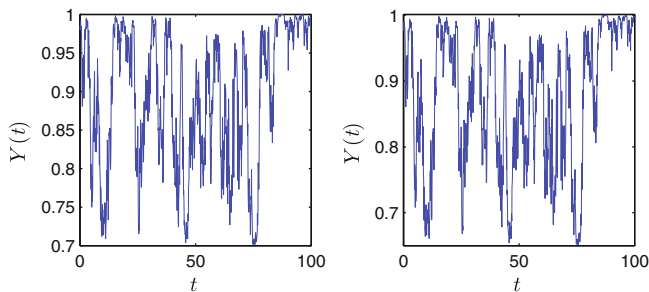
$$\ddot{X}(t) + Y(t)X(t) = 0, \quad t \geq 0, \quad (7.32)$$

where  $Y(t) = g/L(t)$ ,  $g$  is the gravitational constant,  $Y(t) = a + (b - a)\Phi(Z(t))$ ,  $\Phi$  denotes the distribution of  $N(0, 1)$ , and  $Z(t)$  is a stationary Ornstein–Uhlenbeck process with mean 0 and variance 1, that can be approximated by  $Z_{n+1} = \rho Z_n + \sqrt{1 - \rho^2}W_n$ ,  $|\rho| < 1$ , where  $Z_n$  is an approximation for  $Z(n\Delta t)$ ,  $\Delta t > 0$  denotes the time step, and  $W_n \sim \text{iid } N(0, 1)$ . The difference version of (7.32) is of the type in (7.22) with state vector  $X_n = (X(n\Delta t), X((n-1)\Delta t), Z_n)'$ , matrix

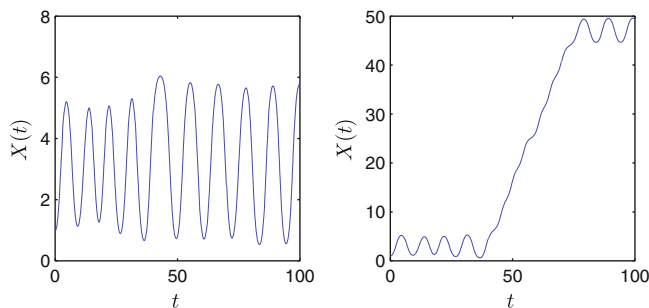
$$A_n(X_n) = \begin{bmatrix} 2X(n\Delta t) - X((n-1)\Delta t) - (\Delta t)^2 X(n\Delta t)(a + (b-a)\Phi(Z_n)) \\ X(n\Delta t) \\ \rho Z_n \end{bmatrix},$$

matrix  $B(X_n) = (0, 0, \sqrt{1 - \rho^2})'$ , and driving noise  $W_n \sim N(0, 1)$ .

Figure 7.4 shows samples of  $Y(t)$  in  $[0, 100]$  for  $\rho = 0.95$ , ( $a = 0.7, b = 1.0$ ) (left panel), and ( $a = 0.65, b = 1.0$ ) (right panel) generated with a time step  $\Delta t = 0.1$ . The corresponding samples of  $X(t)$  with  $X(0) = 1$  are in Fig. 7.5. The pendulum oscillation seem to be stable for ( $a = 0.70, b = 1.0$ ) but are unstable for ( $a = 0.65, b = 1.0$ ).  $\diamond$



**Fig.7.4** Samples of  $Y(t)$  for  $\rho = 0.95$ ,  $(a = 0.7, b = 1.0)$  (left panel) and  $(a = 0.65, b = 1.0)$  (right panel)



**Fig.7.5** Samples of  $X(t)$  for  $\rho = 0.95$ ,  $(a = 0.7, b = 1.0)$  (left panel) and  $(a = 0.65, b = 1.0)$  (right panel)

Figure 7.4 shows that the solutions of equations with random coefficients exhibit a broad range of features depending on the properties of their random coefficients. The characterization of the complex behavior of the solutions of equations with random coefficients is a major challenge that poses notable difficulties in the development of efficient and accurate approximate methods for solving this class of equations.

### 7.3.3 Conditional Analysis

The case of linear difference equations with random coefficients varying in time at random or according to Markov chain has been examined in a previous section. We now examine the case in which the coefficients of these equations follow a semi-Markov process. On samples of these coefficients, the state  $X_n$  satisfies difference equations with deterministic coefficients, so that methods of random vibrations can be applied to find conditional statistics for  $X_n$ .

Suppose matrices  $A_n$  and  $B_n$  in (7.22) change according to semi-Markov process with random transition times  $0 = T_0 < T_1 < \dots < T_r < \dots$  defined by  $T_r =$

$T_{r-1} + S_r, r = 1, 2, \dots$ , where  $\{S_r\}$  are  $\{1, 2, \dots\}$ -valued iid random variables. The recurrence formula (7.22) on samples of  $A_n$  and  $B_n$  defines difference equations with deterministic coefficients driven by random noise in each time interval  $[T_{r-1}(\omega), T_r(\omega))$ .

We consider the special case of (7.22) in which  $A_n(X_n)$  is linear in  $X_n$  and  $B_n(X_n)$  does not depend on  $X_n$ , that is, linear models with additive noise. The driving noise  $W_n$  is white with mean  $\mu_w(n)$  and covariance matrix  $\gamma_w(n)$ . On samples  $(A^{(r)}, B^{(r)})$  of  $\{A_n, B_n\}$  in the time interval  $[T_{r-1}(\omega), T_r(\omega))$ , the state equation is

$$X_{T_{r-1}(\omega)+s+1}^{(r)} = A^{(r)} X_{T_{r-1}(\omega)+s}^{(r)} + B^{(r)} W_{T_{r-1}(\omega)+s}, \quad r = 1, 2, \dots, \quad (7.33)$$

where  $s = 0, 1, \dots, T_r(\omega) - T_{r-1}(\omega)$  is a local temporal coordinate and  $A^{(r)}; B^{(r)}$  are constant deterministic matrices.

**Theorem 7.8** *In the time intervals  $[T_{r-1}(\omega), T_r(\omega))$  on a sample  $(A^{(r)}, B^{(r)})$  of  $(A_n, B_n)$ , the mean vector  $\mu^{(r)}(s) = E[X_{T_{r-1}(\omega)+s}^{(r)}]$  and the covariance matrix  $\gamma^{(r)}(s) = E[(X_{T_{r-1}(\omega)+s}^{(r)} - \mu_r(s))(X_{T_{r-1}(\omega)+s}^{(r)} - \mu_r(s))']$  satisfy the equations*

$$\begin{aligned} \mu^{(r)}(s+1) &= A^{(r)} \mu^{(r)}(s) + B^{(r)} \mu_w(s) \\ \gamma^{(r)}(s+1) &= A^{(r)} \gamma^{(r)}(s) (A^{(r)})' + B^{(r)} \gamma_w(s) (B^{(r)})' \end{aligned} \quad (7.34)$$

for  $s = 0, 1, \dots, T_r(\omega) - T_{r-1}(\omega)$ , where the second moment properties of  $W_n$  are given in the local coordinate.

*Proof* The expectation of (7.33) gives the first formula in (7.34). The second formula results by subtracting the mean equation from (7.33), multiplying the resulting equation with its transpose, and taking the expectation.

The conditional mean and covariance of  $X_n$  in  $[T_0(\omega) = 0, T_1(\omega)]$  can be obtained from (7.34) and second moment properties of  $X_0$ . The mean and covariance of the state at  $T_1(\omega)$  provide initial conditions for (7.34) to continue calculations in  $[T_1(\omega), T_2(\omega)]$  and so on.  $\blacktriangle$

**Theorem 7.9** *Under the conditions in Theorem 7.8, the covariance function  $c(p, q) = E[(X_p - E[X_p])(X_q - E[X_q])']$  of  $X_n$  can be calculated from*

$$\begin{aligned} c(s, s+p) &= \gamma_r(s) ((A^{(r)})^p)' \\ c(s, t) &= \gamma_r(s) ((A^{(r)})^{T_r(\omega)-s})' ((A^{(r+1)})^t)' \end{aligned} \quad (7.35)$$

for  $p \geq 0$ ,  $s, s+p$  local coordinates in  $[T_{r-1}(\omega), T_r(\omega)]$ , and  $t$  local coordinate in  $[T_r(\omega), T_{r+1}(\omega)]$ .

*Proof* We have  $\tilde{X}_{s+p} = (A^{(r)})^p \tilde{X}_s + \text{"noise"}$  in local coordinates, where  $\tilde{X}_s = X_s - E[X_s]$  and "noise" includes all terms involving driving noise centered at its mean. Hence,  $E[\tilde{X}_s \tilde{X}_{s+p}'] = E[\tilde{X}_s ((A^{(r)})^p \tilde{X}_s)']$ , which gives the first formula in (7.35). For example,  $\tilde{X}_{s+1} = A^{(r)} \tilde{X}_s + B^{(r)} \tilde{W}_s$  so that  $E[\tilde{X}_s \tilde{X}_{s+1}'] = \gamma_r(s) (A^{(r)})'$ , where  $\tilde{W}_s = W_s - E[W_s]$ .



For the second formula, note that  $X_t = (A^{(r+1)})^t X_0 + \text{“noise”}$  in  $[T_r(\omega), T_{r+1}(\omega)]$  and  $X_{T_r(\omega)} = (A^{(r)})^{T_r(\omega)-s} X_s + \text{“noise”}$  in  $[T_{r-1}(\omega), T_r(\omega)]$ . Since  $X_0$  for the time interval  $[T_r(\omega), T_{r+1}(\omega)]$  is  $X_{T_r(\omega)}$  at the end of  $[T_{r-1}(\omega), T_r(\omega)]$ , we can calculate  $X_t$  from  $X_t = (A^{(r+1)})^t (A^{(r)})^{T_r(\omega)-s} X_s + \text{“noise”}$ . The expectation  $E[\tilde{X}_s \tilde{X}_t']$  gives the stated formula. Similar arguments can be used to find the covariance function at arbitrary times. ▲

Unconditional second moment properties of  $X_n$  can be obtained by Monte Carlo simulation and the conditional second moment properties given by (7.34)–(7.35). A similar approach can be followed for other properties of  $X_n$ .

*Example 7.17* Let  $X_n$  be a real-valued process defined by  $X_{n+1} = Y_n X_n + W_n$ ,  $n = 0, 1, \dots$ , that is a special case of (7.22) with  $d = 1$ ,  $A_n(X_n) = Y_n X_n$ ,  $B_n = 1$ , and  $W_n$  independent  $N(0, \sigma^2)$  variables. The times  $S_r = T_r - T_{r-1}$ ,  $r = 1, 2, \dots$ , between consecutive jumps of the semi-Markov process  $Y_n$  are assumed to be iid  $\{1, 2, \dots\}$ -valued random variables following a geometrical distribution, that is,  $P(S_r = k) = (1 - p)^{k-1} p$ , where  $k = 1, 2, \dots$  and  $0 < p < 1$ . The mean and variance of  $S_k$  are  $E[S_k] = 1/p$  and  $\text{Var}[S_k] = (1 - p)/p^2$ . We examine in detail time continuous versions of this model in a subsequent section. ◇

Consider the special case of (7.22) defining a linear model with additive noise, that is,  $X_{n+1} = A X_n + B W_n$ , where  $A_n = A$  and  $B_n X_n = B$  are time invariant. The second moment properties of the conditional state  $X_n | (A, B)$  can be obtained from the theory of discrete linear systems in Sect. 7.2.1. If the driving noise is Gaussian,  $X_n | (A, B)$  is a Gaussian sequence. The probability law of the unconditional series  $X_n$  can be obtained by eliminating the condition on  $(A, B)$ . The series  $\{X_n\}$  is not Gaussian.

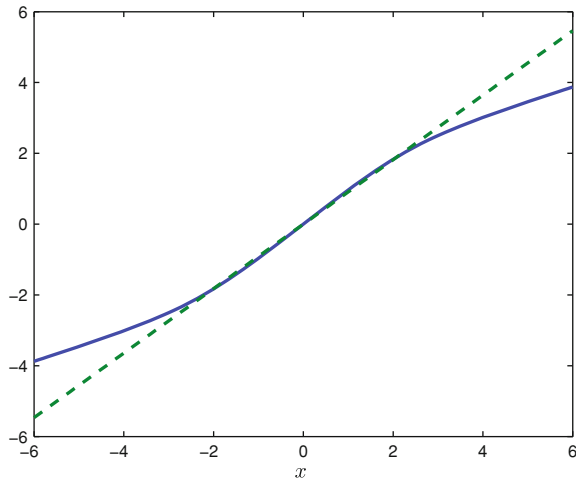
*Example 7.18* Suppose the real-valued series  $X_n$  is the state of a linear model with additive noise, initial state  $X_0 \sim N(0, 1)$ , coefficients  $A \sim U(a_1, a_2)$ ,  $-1 < a_1 < a_2 < 1$ , and  $B \sim U(b_1, b_2)$ ,  $0 < b_1 < b_2 < \infty$ , independent of each other, and driving noise  $W_n \sim N(0, 1)$ . The mean  $\tilde{\mu}_n$ , variance  $\tilde{\gamma}_n$ , and covariance function  $\tilde{c}(n, n + q)$  of the conditional series  $X_n | (A, B)$  satisfy the equations

$$\begin{aligned}\tilde{\mu}_{n+1} &= A \tilde{\mu}_n, \\ \tilde{\gamma}_{n+1} &= A^2 \tilde{\gamma}_n + B^2, \\ \tilde{c}(n, n + q) &= A^q \tilde{\gamma}_n, \quad q \geq 0,\end{aligned}\tag{7.36}$$

implying  $\tilde{\mu}_n = 0$ ,  $\lim_{n \rightarrow \infty} \tilde{\gamma}_n = \tilde{\gamma} = B^2/(1 - A^2)$ ,  $\lim_{m, n \rightarrow \infty} \tilde{c}(m, n) = \tilde{\gamma} A^{|m-n|}$ , and  $P(X_n \leq x | A, B) = \Phi(x/\sqrt{\tilde{\gamma}_n})$ . Properties of  $X_n$  result by eliminating the condition on  $(A, B)$ . For example,  $P(X_n \leq x) = E[P(X_n \leq x | A, B)]$  and  $\lim_{n \rightarrow \infty} P(X_n \leq x) = E[\Phi(x/\sqrt{\tilde{\gamma}})]$ . Figure. 7.6 shows the variation of  $\Phi^{-1}(P(X_n \leq x))$  with  $x$  for  $n \rightarrow \infty$ . Since the solid line is not straight, the asymptotic marginal distribution of  $X_n$  cannot be Gaussian. ◇

*Proof* The expectation of  $X_{n+1} = A X_n + B W_n$  conditional on  $(A, B)$  gives the conditional mean equation. Since  $\tilde{\mu}_0 = E[X_0] = 0$ , we have  $\tilde{\mu}_n = 0$ . The recurrence

**Fig. 7.6** Dependence of  $P(X_n \leq x) = E[P(X_n \leq x | A, B)]$  on  $x$  for  $n \rightarrow \infty$



formula for  $\tilde{\gamma}_n$  follows by calculating the expectation of the square of  $X_{n+1} = AX_n + BW_n$  conditional on  $(A, B)$ . Repeated application of this formula gives  $\tilde{\gamma}_n = A^{2n} + B^2 \sum_{k=0}^{n-1} A^{2(n-k-1)}$  so that  $\lim_{n \rightarrow \infty} \tilde{\gamma}_n = B^2/(1 - A^2)$  since  $|A| < 1$  a.s. by assumption and  $\sum_{k=0}^{n-1} A^{2(n-k-1)}$  sums to  $(1 - A^{2n})/(1 - A^2)$ . The conditional covariance equation follows by direct calculations from  $\tilde{c}(n, n+1) = E[X_n X_{n+1} | A, B] = E[X_n (AX_n + BW_n) | A, B] = A\tilde{\gamma}_n$ ,  $\tilde{c}(n, n+2) = E[X_n X_{n+2} | A, B] = AE[X_n X_{n+1} | A, B] = A^2\tilde{\gamma}_n$ , and so on. The asymptotic marginal distribution of the unconditional series  $X_n$  results from

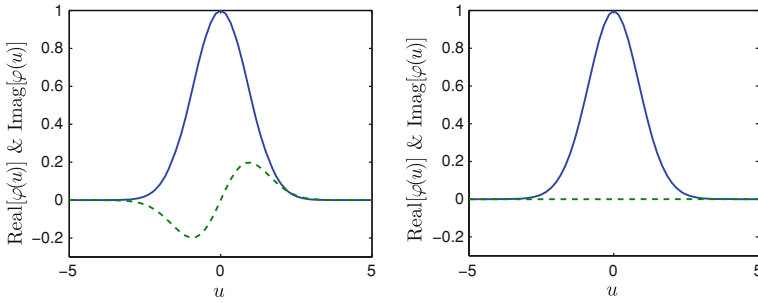
$$\lim_{n \rightarrow \infty} P(X_n \leq x) = \lim_{n \rightarrow \infty} E \left[ \Phi \left( \frac{x}{\sqrt{\tilde{\gamma}_n}} \right) \right] = E \left[ \lim_{n \rightarrow \infty} \Phi \left( \frac{x}{\sqrt{\tilde{\gamma}_n}} \right) \right] = E \left[ \Phi \left( \frac{x\sqrt{1-A^2}}{B} \right) \right],$$

where the last two equalities hold by bounded convergence and the continuity of the Gaussian distribution.

Regarding the plot in Fig. 7.6, if the asymptotic distribution  $P(X_n \leq x)$  were Gaussian with mean  $\mu$  and standard deviation  $\sigma$ , then  $\Phi^{-1}(P(X_n \leq x))$  should be equal to  $(x - \mu)/\sigma$ , so that it should plot as a straight line against  $x$ . ▲

**Example 7.19** Consider the linear model in (7.25). The characteristic function  $\varphi_n(u) = E[\exp(iuX_n)]$  of  $X_n$  can be obtained from  $\varphi_{n+1}(u) = E[\varphi_n(uA_n)]\varphi_W(u)$ , where  $\varphi_W(u) = E[\exp(iuW_n)]$ . Figure 7.7 shows with solid and dotted lines the real and the imaginary parts of  $\varphi_n(u)$  for  $n = 1$  (left panel) and  $n = 10$  (right panel) for  $X_0 = 1$ ,  $A_n \sim U(a_1, a_2)$  with  $a_1 = -0.2$  and  $a_2 = 0.9$ , and  $\gamma_W = 1$ . The characteristic function  $\varphi_n(u)$  is real-valued and time invariant for  $n \geq 10$  indicating that for relatively large times the density of  $X_n$  is an even function.

Recurrence formulas for the characteristic function are impractical for vector time series since their implementation involves numerical evaluation of multidimensional integrals at each step  $n$ . Also, the determination of the stationary marginal charac-



**Fig. 7.7** Real and imaginary parts of  $\varphi_n(u)$  for  $n = 1$  (left panel) and  $n = 10$  (right panel)

teristic  $\varphi(u)$  of  $X_n$ , provided it exists, is rather difficult since  $\varphi(u)$  is the solution of  $\varphi(u) = E[\varphi(uA_n)]\varphi_W(u)$ .  $\diamond$

*Proof* The expectation of  $\exp(iuX_{n+1})$  is

$$\varphi_{n+1}(u) = E[\exp(iuA_nX_n)]E[\exp(iuW_n)] = E[\varphi_n(uA_n)]\varphi_W(u)$$

by properties of  $W_n$  and of the conditional expectation. For the special case in which  $A_n = \rho \in (-1, 1)$  is deterministic and  $W_n \sim N(0, \gamma_w)$ , the stationary characteristic function of  $X_n$  can be obtained simply, and is  $\varphi(u) = \exp(-\gamma_w u^2 / (2(1 - \rho^2)))$ .  $\blacktriangle$

### 7.3.4 Stochastic Reduced Order Models

Stochastic reduced order models are discussed in Sects. A.3 and A.4. These models are simple random elements whose samples are selected from samples of target random elements and, generally, are not equally likely. For example, a SROM  $\tilde{A}_n$  for  $A_n$  in (7.25) is a simple random variable with samples  $(\tilde{a}_1, \dots, \tilde{a}_m)$  that have probabilities  $(p_1, \dots, p_m)$ . We calculate properties of the state  $X_n$  in (7.25) from those of the state of this equation with  $\tilde{A}_n$  in place of  $A_n$  under the assumption that  $\{A_n\}$  is an iid series.

**Theorem 7.10** *The marginal distribution  $\tilde{F}_n$  and the moments  $\tilde{\mu}_n(r) = E[\tilde{X}_n^r]$  of order  $r \geq 1$  of  $\tilde{X}_n$  defined by (7.25) with  $\tilde{A}_n$  in place of  $A_n$  satisfy the recurrence formulas*

$$\begin{aligned} \tilde{F}_{n+1}(x) &= \sum_{j=1}^m p_j \int \tilde{F}_n((x-u)/\tilde{a}_j) dF_w(u) du \quad \text{and} \\ \tilde{\mu}_{n+1}(r) &= \sum_{j=1}^m p_j \sum_{s=0}^r \frac{r!}{s!(r-s)!} \tilde{a}_j^s \tilde{\mu}_n(s) E[W_n^{r-s}], \quad n = 0, 1, \dots, \end{aligned} \quad (7.37)$$

where  $\tilde{F}_0(x) = F_0(x) = P(X_0 \leq x)$  and  $\tilde{\mu}_0(r) = \mu_0(r) = E[X_0^r]$ . It is assumed that  $\{A_n\}$  is an iid series.

*Proof* The distribution of  $\tilde{X}_{n+1}$  defined by (7.25) with  $\tilde{A}_n$  in place of  $A_n$  can be calculated from

$$\begin{aligned}\tilde{F}_{n+1}(x) &= E[1(\tilde{X}_{n+1} \leq x)] = E\{E[1(\tilde{A}_n \tilde{X}_n + W_n \leq x) \mid \tilde{A}_n]\} \\ &= E\left\{\int \tilde{F}_n((x-u)/\tilde{A}_n) dF_w(u)\right\} = \sum_{j=1}^m p_j \int \tilde{F}_n((x-u)/\tilde{a}_j) dF_w(u) du,\end{aligned}$$

where  $F_w$  denotes the distribution of  $W_n$ . The recurrence formula for moments is

$$\begin{aligned}E[\tilde{X}_{n+1}^r \mid \tilde{A}_n] &= E\{E[(\tilde{A}_n \tilde{X}_n + W_n)^r \mid \tilde{A}_n]\} = E\left\{\sum_{s=0}^r \frac{r!}{s!(r-s)!} \tilde{A}_n^s E[\tilde{X}_n^s] E[W_n^{r-s}]\right\} \\ &= \sum_{j=1}^m p_j \sum_{s=0}^r \frac{r!}{s!(r-s)!} \tilde{a}_j^s E[\tilde{X}_n^s] E[W_n^{r-s}],\end{aligned}$$

where  $r \geq 0$  is an integer.  $\blacktriangle$

*Example 7.20* Let  $X_n$  be define by (7.25) with  $A_n$  independent  $U(a_1, a_2)$ ,  $-1 < a_1 < a_2 < 1$ , random variables,  $W_n \sim N(0, 1)$  a Gaussian white noise, and  $X_0 = 1$ . Let  $\tilde{A}_n$  be a SROM for  $A_n$  with  $m \geq 1$  samples  $\tilde{a}_j = a_1 + (j-1/2)(a_2 - a_1)/m$  and probabilities  $p_j = 1/m$ ,  $j = 1, \dots, m$ . Figure 7.8 shows with solid and dotted lines Monte Carlo estimates and SROM-based approximations of the first six moments of  $X_n$  for  $n = 0, 1, \dots, 10$ , that is the moments  $E[X_n^r]$ ,  $r = 1, \dots, 6$ , during the time interval  $[0, 10]$ . The Monte Carlo estimates are based on 100000 independent samples of  $X_n$  and a SROM  $\tilde{A}_n$  with  $m = 5$  samples. The plots are for  $(a_1, a_2) = (-0.2, 0.9)$ . The largest discrepancy in  $[0, 10]$  between Monte Carlo estimates and SROM-based approximations for  $E[X_n^r]$  are 0.8475%, 1.5628%, 3.5342%, 3.5174 %, 5.1714%, and 6.5156% for  $r = 1, \dots, 6$ . The second recurrence formula in (7.37) has been used to obtained SROM-based moments.  $\diamond$

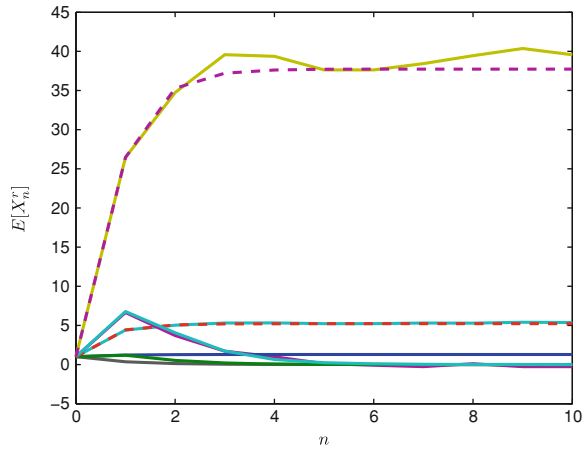
**Theorem 7.11** *The average discrepancy  $\varepsilon(n) = E[|X_n - \tilde{X}_n|]$  between the exact and the SROM-based solution satisfies the relationship*

$$\varepsilon(n+1) \leq \alpha \sum_{i=0}^n \beta^i, \quad (7.38)$$

where  $\alpha = \max_n \{E[|A_n - \tilde{A}_n|]E[|X_n|]\}$  and  $\beta = \max_n \{E[|\tilde{A}_n|]\}$ .

*Proof* We have  $|X_{n+1} - \tilde{X}_{n+1}| = |A_n X_n - \tilde{A}_n \tilde{X}_n| \leq |A_n - \tilde{A}_n| |X_n| + |\tilde{A}_n| |X_n - \tilde{X}_n|$  implying  $\varepsilon(n+1) \leq E[|A_n - \tilde{A}_n|]E[|X_n|] + E[|\tilde{A}_n|]\varepsilon(n)$ . Since  $\varepsilon(0) = 0$ , we have  $\varepsilon(n+1) \leq \alpha + \beta\varepsilon(n) \leq \alpha + \beta(\alpha + \beta\varepsilon(n-1)) \leq \alpha \sum_{i=0}^n \beta^i$ . If  $\alpha$  and  $\beta$  are finite and  $\beta < 1$ , then  $\varepsilon(n) \leq \alpha(1 - \beta^n)/(1 - \beta) \leq \alpha/(1 - \beta)$  at all times.  $\blacktriangle$

**Fig. 7.8** Monte Carlo estimates and SROM-based approximations of  $E[X_n^r]$ ,  $r = 1, \dots, 6$  for  $(a_1, a_2) = (-0.2, 0.9)$



### 7.3.5 Stochastic Galerkin and Collocation Methods

The first step of the stochastic Galerkin and collocation methods is the representation of the random elements in the definition of a stochastic equation by deterministic functions that depend on small numbers of random variables. These methods are discussed extensively later in this chapter in the context of stochastic differential equations with random coefficients and, also, in the following two chapters. Their usefulness to stochastic equations of the type considered here seems to be rather limited.

For example, the state in the recurrence formula (7.25) has the expression

$$X_{n+1} = \left( \prod_{i=0}^n A_i \right) X_0 + \sum_{i=0}^n W_i \prod_{j=i+1}^{n-1} A_j + W_n, \quad n = 0, 1, \dots, \quad (7.39)$$

so that it is a function of  $(A_0, A_1, \dots, A_n)$  in addition to the driving noise and initial state. If  $\{X_n\}$  become stationary as  $n \rightarrow \infty$  and we are interested in properties of the stationary version of  $\{X_n\}$ , the random vector  $(A_0, A_1, \dots, A_n)$  will have an infinite number of coordinates. Also,  $(A_0, A_1, \dots, A_n)$  can be a very large vector if the behavior of  $\{X_n\}$  over relatively large time intervals is of interest, so that direct solutions by the Galerkin and collocation methods are likely to be impractical.

### 7.3.6 Taylor Series

We have seen that the real-valued series defined by (7.39) is the solution of the finite difference equation in (7.25). The formula in (7.39) shows that  $X_{n+1}$  is a function of  $(A_0, A_1, \dots, A_n)$ ,  $X_0$ , and  $(W_0, W_1, \dots, W_n)$ . We view  $X_{n+1}$  as a function of  $(A_0, A_1, \dots, A_n)$ , expand it in Taylor series about  $(E[A_0], E[A_1], \dots, E[A_n])$ , and

use a truncated version of this series to calculate moments of  $X_{n+1}$  approximately. It is assumed that  $\{A_n\}$  and  $\{W_n\}$  are iid series.

**Theorem 7.12** *The approximate mean and variance of  $X_{n+1}$  given by its first order Taylor expansion about the mean of  $(A_0, A_1, \dots, A_n)$  are*

$$\begin{aligned} E[X_{n+1}] &\simeq E[X_0]\rho^{n+1} \\ \text{Var}[X_{n+1}] &\simeq \text{Var}[X_0]\rho^{2(n+1)} + \gamma_w \sum_{i=0}^n \rho^{2(n-i)} \\ &\quad + \gamma_a \sum_{k=0}^n \left( E[A_0^2]\rho^{2n} + \gamma_w \sum_{i=0}^{k-1} \rho^{2(n-i-1)} \right) \end{aligned} \quad (7.40)$$

with the convention  $\sum_{i=0}^{-1}(\cdot) = 0$  and the notations  $\rho = E[A_n]$ ,  $\gamma_a = \text{Var}[A_n]$ , and  $\gamma_w = \text{Var}[W_n]$ .

*Proof* The random variables  $X_{n+1}$  and  $\partial X_{n+1}/\partial A_k$  with  $(A_0, A_1, \dots, A_n)$  set equal to its mean value are

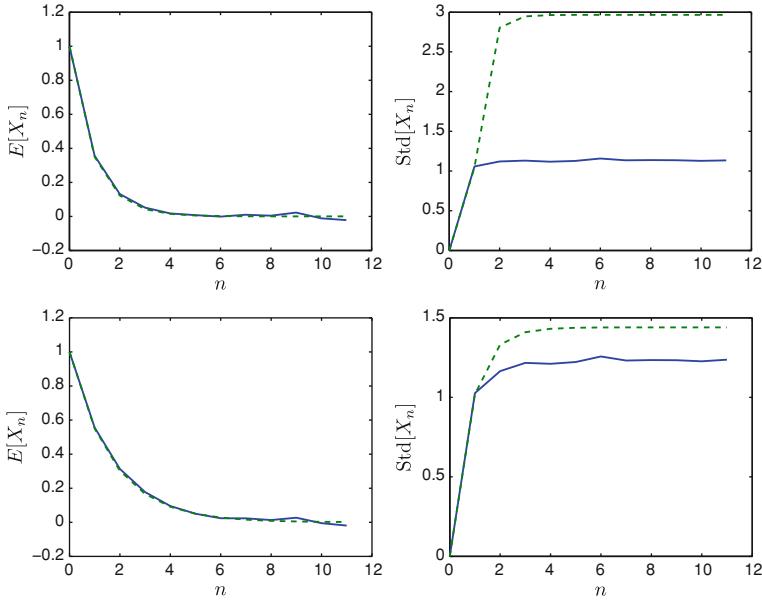
$$X_0\rho^{n+1} + \sum_{i=0}^{n-1} W_i\rho^{n-i} + W_n \quad \text{and} \quad X_0\rho^n + \sum_{i=0}^{k-1} W_i\rho^{n-i-1},$$

respectively, so that the first order Taylor approximation of  $X_{n+1}$  about the mean of  $(A_0, A_1, \dots, A_n)$  is

$$X_{n+1} \simeq \left( X_0\rho^{n+1} + \sum_{i=0}^{n-1} W_i\rho^{n-i} + W_n \right) + \sum_{k=0}^n \left( X_0\rho^n + \sum_{i=0}^{k-1} W_i\rho^{n-i-1} \right) (A_k - E[A_k]). \quad (7.41)$$

The mean and variance of this approximate representation for  $X_{n+1}$  are given by (7.40). The calculation of these moments use the assumptions that  $(A_0, A_1, \dots)$  and  $(W_0, W_1, \dots)$  are mutually independent, and are also independent of  $X_0$ .  $\blacktriangle$

*Example 7.21* Suppose the initial state is deterministic and equal to  $X_0 = 1$  so that the approximate mean and variance of the state can be calculated from (7.40) with  $E[X_0] = 1$  and  $\text{Var}[X_0] = 0$ . The plots in left and the right panels of Fig. 7.9 are means and standard deviations of  $X_n$  as functions of  $n$  for  $A_n \sim U(a_1, a_2)$  and  $\gamma_w = 1$ . The solid and dotted lines are Monte Carlo estimates based on 10000 samples and first order Taylor approximations. The top and bottom panels are for  $(a_1, a_2) = (-0.2, 0.9)$  and  $(a_1, a_2) = (0.2, 0.9)$ , respectively. The approximate means are satisfactory but the approximate standard deviations are inaccurate for  $(a_1, a_2) = (-0.2, 0.9)$ . The accuracy of the approximate standard deviations deteriorates as the uncertainty in  $A_n$  increases from  $A_n \sim U(0.2, 0.9)$  to  $A_n \sim U(-0.2, 0.9)$ . Numerical values for the plots in Fig. 7.9 have been obtained by using the variance formula in (7.40).  $\diamond$



**Fig. 7.9** Monte Carlo estimates (solid lines) and Taylor approximations (dotted lines) of  $E[X_n]$  and  $\text{Std}[X_n]$ . The top and bottom panels are for  $A_n \sim U(-0.2, 0.9)$  and  $A_n \sim U(0.2, 0.9)$ , respectively. The left and right panels show means and standard deviations of  $X_n$

### 7.3.7 Perturbation Series

The perturbation method is useful for analyzing dynamic systems with small nonlinearities. We apply this method to stochastic difference equations with random coefficients that have small uncertainty.

Consider the model in (7.25) with  $E[X_0] = 0$ ,  $E[X_0^2] < \infty$ , random coefficients  $A_n = \rho + \varepsilon Y_n$  depending on the iid random variables  $\{Y_n\}$  with mean 0 and variance  $\gamma_y$  such that  $|A_n| < 1$  a.s., and  $|\varepsilon| \ll 1$ . Assume that the power series representation,

$$X_n = X_{n,0} + \varepsilon X_{n,1} + \varepsilon^2 X_{n,2} + \cdots, \quad (7.42)$$

of the solution of (7.25) is convergent, and approximate  $X_n$  by the first two terms of this series, that is,  $X_n \simeq \tilde{X}_n = X_{n,0} + \varepsilon X_{n,1}$ . The time series  $(X_{n,0}, X_{n,1})$  satisfy the difference equations

$$\begin{aligned} X_{n+1,0} &= \rho X_{n,0} + W_n \quad \text{and} \\ X_{n+1,1} &= \rho X_{n,1} + Y_n X_{n,0}. \end{aligned} \quad (7.43)$$

Note that these equations have the same operator but different inputs.

**Theorem 7.13** *The variance and covariance functions of  $X_{n,0}$ ,  $X_{n,1}$ , and  $\tilde{X}_n = X_{n,0} + \varepsilon X_{n,1}$  in (7.37) can be calculated from*

$$\begin{aligned}\gamma_{00}(n+1) &= \rho^2 \gamma_{00}(n) + \gamma_w \\ \gamma_{11}(n+1) &= \rho^2 \gamma_{11}(n) + \gamma_y \gamma_{00}(n) \\ \gamma_{01}(n+1) &= \rho^2 \gamma_{01}(n) \\ c_{kl}(m, n) &= \rho^{|m-n|} \gamma_{kl}(m \wedge n), \quad k, l = 0, 1,\end{aligned}\tag{7.44}$$

where  $\gamma_{kl}(n) = E[X_{n,k}X_{n,l}]$  and  $c_{kl}(m, n) = E[X_{m,k}X_{n,l}]$ . The sequences  $X_{n,0}$ ,  $X_{n,1}$ , and  $\tilde{X}_n$  have mean 0.

*Proof* The bivariate vector  $Z_n = (X_{n,0}, X_{n,1})'$  satisfies the difference equation

$$\begin{bmatrix} X_{n+1,0} \\ X_{n+1,1} \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ Y_n & \rho \end{bmatrix} \begin{bmatrix} X_{n,0} \\ X_{n,1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} W_n = \alpha_n Z_n + \beta W_n,$$

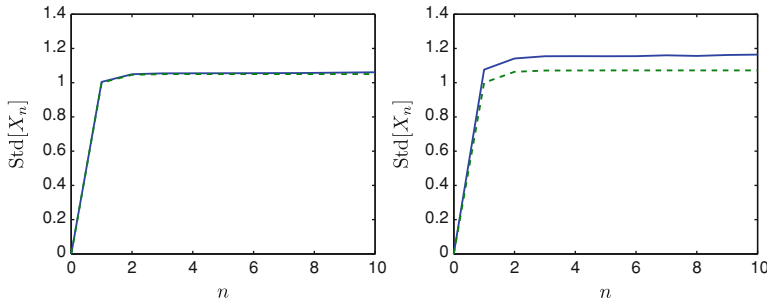
so that  $E[Z_{n+1}] = E[\alpha_n Z_n] = E\{E[\alpha_n Z_n | Y_n]\} = E\{\alpha_n E[Z_n]\} = E[\alpha_n]E[Z_n]$  since  $\alpha_n$  is a function of  $Y_n$ ,  $Z_n$  depends on  $(Y_{n-1}, Y_{n-2}, \dots)$ ,  $\{Y_n\}$  is an iid sequence, and  $E[W_n] = 0$  by assumption. The conditional covariance  $\tilde{\gamma}(n) = E[Z_n Z_n' | Y_n]$  is the solution of  $\tilde{\gamma}(n+1) = \alpha_n \tilde{\gamma}(n) \alpha_n' + \beta \beta' \gamma_w$  (Sect. 7.2.1), so that the unconditional covariance  $\gamma(n) = E[Z_n Z_n']$  of  $Z_n$  can be obtained from

$$\begin{aligned}\gamma(n+1) &= E[(\alpha_n Z_n + \beta W_n)(\alpha_n Z_n + \beta W_n)'] = E[\alpha_n Z_n Z_n' \alpha_n'] + \beta \beta' \\ &= E\{E[\alpha_n Z_n Z_n' \alpha_n' | Y_n, Y_{n-1}, Y_{n-2}, \dots]\} \\ &= E\{\alpha_n E[Z_n Z_n' | Y_{n-1}, Y_{n-2}, \dots] \alpha_n'\} + \beta \beta' = E\{\alpha_n \gamma(n) \alpha_n'\} + \beta \beta'\end{aligned}$$

since  $\alpha_n$  and  $Z_n$  are independent of  $W_n$ ,  $W_n$  has mean 0 and variance 1,  $\alpha_n$  is a function of only  $Y_n$  and  $Z_n$  depends on  $(Y_{n-1}, Y_{n-2}, \dots)$ , and  $\{Y_n\}$  is an iid series. This recurrence formula gives the first three equations in (7.44). Similar arguments can be used to obtain the last equality in (7.44).  $\blacktriangle$

**Example 7.22** Suppose  $Y_n$  and  $W_n$  in Theorem 7.13 are iid  $U(-1, 1)$  and iid  $N(0, 1)$ , respectively. Under the assumption that the initial state  $X_0$  is deterministic, we have  $X_{0,0} = 0$ ,  $X_{0,1} = 0$ ,  $E[X_{n,0} + \varepsilon X_{n,1}] = 0$ ,  $\gamma_{kl}(0) = 0$ ,  $E[(X_{n,0} + \varepsilon X_{n,1})^2] = \gamma_{00}(n) + \varepsilon^2 \gamma_{11}(n)$ . The solid and dotted lines in Fig. 7.10 are Monte Carlo estimates and first order perturbation solutions for the standard deviation  $\text{Std}[X_n]$  of  $X_n$  corresponding to  $(\rho = 0.3, \varepsilon = 0.2)$  (left panel) and  $(\rho = 0.3, \varepsilon = 0.69)$  (right panel). The Monte Carlo estimates are based on 10000 independent samples of  $X_n$ . The accuracy of perturbation solutions decreases with  $\varepsilon$  but remains satisfactory for the cases considered here.  $\diamond$





**Fig. 7.10** Monte Carlo estimates (*solid lines*) and perturbation approximations (*dotted lines*) of  $\text{Std}[X_n]$  for  $(\rho = 0.3, \varepsilon = 0.2)$  (*left panel*) and  $(\rho = 0.3, \varepsilon = 0.69)$  (*right panel*)

## 7.4 Stochastic Differential Equations with Random Coefficients

We consider equations of the type in (7.1) and (7.2), state conditions for the existence of unique solutions for these equations, present methods for solving differential equations with random coefficients and input, and illustrate numerically the implementation of some these methods.

### 7.4.1 General Considerations

The probability laws of the solutions of stochastic differential equations (SDEs) are determined by properties of their drift/diffusion coefficients and driving noise processes, and can be affected significantly by the uncertainty in their coefficients. For example, the solutions of linear SDEs with Gaussian noise are Gaussian processes, but become non-Gaussian if their coefficients are uncertain.

*Example 7.23* Let  $X(t)$ ,  $t \geq 0$ , be the solution of  $\dot{X}(t) + AX(t) = Y(t)$ , where  $Y(t)dt = dB(t)$ ,  $B$  denotes a standard Brownian motion,  $A \geq a > 0$  a.s. is a random variable,  $E[X(0)] = \mu_0$ , and  $\text{Var}[X(0)] = \gamma_0$ . The conditional mean  $\mu(t; A) = E[X(t) | A]$  and variance  $\gamma(t; A) = \text{Var}[X(t) | A]$  satisfy the equations

$$\begin{aligned}\dot{\mu}(t; A) &= -A\mu(t; A) \\ \dot{\gamma}(t; A) &= -2A\gamma(t; A) + 1\end{aligned}$$

so that  $\mu(t; A) = \mu_0 \exp(-At)$  and  $\gamma(t; A) = \gamma_0 \exp(-2At) + (1 - \exp(-2At))/(2A)$  implying the a.s. convergence  $\mu(t; A) \rightarrow 0$  and  $\gamma(t; A) \rightarrow 1/(2A)$  as  $t \rightarrow \infty$ .

The conditional random variable  $X(t) | A$  is Gaussian with mean  $\mu(t; A)$  and variance  $\gamma(t; A)$ . On the other hand,  $X(t)$  is not Gaussian. For example, estimates of the kurtosis coefficient of  $X(t)$  for  $A \sim U(0.1, 3)$  based on 10000 independent samples are 0.5442, 1.8578, and 2.3515 at times  $t = 1, 5$ , and 10.  $\diamond$

Solutions for stochastic differential equations of the type discussed in Sect. 5.5.1, that is, ordinary differential equations with deterministic coefficients and white noise

input can be defined in the strong and weak sense. We consider here an alternative type of weak solutions for ordinary differential equations with random coefficients that relates to weak convergence in Hilbert spaces. Weak solutions of this type are used extensively to solve stochastic partial differential equations (Sect. 9.4). Following is a brief discussion on weak solutions for ordinary differential equations with random coefficients and input.

The weak formulation for initial/boundary value problems requires to find  $u \in H_1$  such that  $\mathcal{B}(u, v) = \mathcal{J}(v)$ ,  $\forall v \in H_2$ , where  $\mathcal{B} : H_1 \times H_2 \rightarrow \mathbb{R}$  and  $\mathcal{J} : H_2 \rightarrow \mathbb{R}$  are bilinear and linear functionals and  $H_1$  and  $H_2$  denote Hilbert spaces. The Lax-Milgram theorem states conditions that  $\mathcal{B}$  and  $\mathcal{J}$  must satisfy such that  $\mathcal{B}(u, v) = \mathcal{J}(v)$  has a unique solution for the special case  $H_1 = H_2$ . The Babuška-Lax-Milgram theorem [40] removes the restriction  $H_1 = H_2$ .

**Theorem 7.14** (Babuška-Lax-Milgram theorem) *Let  $\mathcal{B} : H_1 \times H_2 \rightarrow \mathbb{R}$  be a continuous bilinear functional. If  $\mathcal{B}$  is weakly coercive, that is, there exists a constant  $c > 0$  such that*

$$\begin{aligned} \sup_{\|v\|_{H_2}=1} |\mathcal{B}(u, v)| &\geq c\|u\|_{H_1} \quad \text{and} \\ \sup_{u \in H_1} |\mathcal{B}(u, v)| &> 0, \quad \forall v \in H_2 \setminus \{0\}, \end{aligned} \quad (7.45)$$

*then for all  $f \in H_2$  there exists a unique solution  $u_f \in H_1$  such that  $\mathcal{B}(u_f, v) = \mathcal{J}(v)$  for all  $v \in H_2$ , where  $\mathcal{J} : H_2 \rightarrow \mathbb{R}$  is given by  $\mathcal{J}(v) = \langle f, v \rangle_{H_2}$ . Moreover, the unique solution  $u_f$  is bounded by  $\|u_f\|_{H_1} \leq \|f\|_{H_2}/c$ .*

**Example 7.24** Let  $X(t)$ ,  $0 \leq t \leq \tau$ , be a real-valued stochastic process defined by the stochastic differential equation  $\dot{X}(t) + AX(t) = Y(t)$ ,  $t \in I = [0, \tau]$ , with initial state  $X(0) = 0$ . It is assumed that the random variable  $A$  and the real-valued stochastic process  $Y(t)$  are defined on the same probability space  $(\Omega, \mathcal{F}, P)$ ,  $A$  is independent of  $Y(t)$ ,  $P(a_1 < A < a_2) = 1$ ,  $0 < a_1 < a_2 < \infty$ , and  $Y(t)$  has finite variance.

Consider the Hilbert spaces

$$\begin{aligned} \mathcal{V} &= \{Z : I \times \Omega \rightarrow \mathbb{R}, Z \in L^2(I \times \Omega), \mathcal{B}(I) \times \mathcal{F} - \text{measurable}\} \quad \text{and} \\ \mathcal{W} &= \{U : I \times \Omega \rightarrow \mathbb{R}, \dot{U} \in L^2(I \times \Omega), \mathcal{B}(I) \times \mathcal{F} - \text{measurable}, U(0) = 0 \text{ a.s.}\} \end{aligned} \quad (7.46)$$

with the inner products

$$\begin{aligned} \langle Z_1, Z_2 \rangle_{\mathcal{V}} &= E \left[ \int_I Z_1(t) Z_2(t) dt \right] = \int_{I \times \Omega} Z_1(t, \omega) Z_2(t, \omega) dt P(d\omega) \quad \text{and} \\ \langle U_1, U_2 \rangle_{\mathcal{W}} &= \langle \dot{U}_1, \dot{U}_2 \rangle_{\mathcal{V}} + \langle U_1, U_2 \rangle_{\mathcal{V}}, \end{aligned} \quad (7.47)$$

respectively (Exercise 7.13), so that  $\mathcal{V} = L^2(I \times \Omega, \mathcal{B}(I) \times \mathcal{F}, \lambda \times P)$ , where  $\lambda(dt) = dt$  denotes the Lebesgue measure on the real line. The norms induced by

these inner products on  $\mathcal{V}$  and  $\mathcal{W}$  are  $\|Z\|_{\mathcal{V}}^2 = \langle Z, Z \rangle_{\mathcal{V}} = E\left[\int_I Z(t)^2 dt\right]$  and  $\|U\|_{\mathcal{W}}^2 = \langle U, U \rangle_{\mathcal{W}} = E\left[\int_I (\dot{U}(t)^2 + U(t)^2) dt\right] = \|U\|_{\mathcal{V}}^2 + \|\dot{U}\|_{\mathcal{V}}^2$ .

Our objective is to find  $X \in \mathcal{W}$  such that  $\mathcal{B}(X, Z) = \mathcal{J}(Z)$ ,  $\forall Z \in \mathcal{V}$ , where the functionals  $\mathcal{B} : \mathcal{W} \times \mathcal{V} \rightarrow \mathbb{R}$  and  $\mathcal{J} : \mathcal{V} \rightarrow \mathbb{R}$  are defined by

$$\begin{aligned}\mathcal{B}(X, Z) &= E\left[\int_I (\dot{X}(t) + AX(t))Z(t)dt\right] \\ &= \int_{I \times \Omega} (\dot{X}(t, \omega) + A(\omega)X(t, \omega))Z(t, \omega)dt P(d\omega) \quad \text{and} \\ \mathcal{J}(Z) &= E\left[\int_I Y(t)Z(t)dt\right] = \int_{I \times \Omega} Y(t, \omega)Z(t, \omega)dt P(d\omega),\end{aligned}\quad (7.48)$$

that is, the integrals of the left and the right sides of  $\dot{X}(t) + AX(t) = Y(t)$  multiplied by  $Z \in \mathcal{V}$  calculated over  $I \times \Omega$  under the measure  $\lambda \times P$ . The functional  $\mathcal{J}$  is linear and continuous, and therefore bounded, and  $\mathcal{B}$  is bilinear, bounded, and weakly coercive. According to Theorem 7.14,  $\dot{X}(t) + AX(t) = Y(t)$  admits a unique solution  $X_Y \in \mathcal{W}$  such that  $\mathcal{B}(X_Y, Z) = \mathcal{J}(Z)$ ,  $\forall Z \in \mathcal{V}$ .  $\diamond$

*Proof* The set  $\mathcal{V}$  is a Hilbert space since it coincides with  $L^2(I \times \Omega, \mathcal{B}(I) \times \mathcal{F}, \lambda \times P)$ . Since  $U \in \mathcal{W}$  implies that  $U$  has finite variance,  $\mathcal{W}$  is a subset of  $\mathcal{V}$ . Moreover,  $\mathcal{W}$  is a linear subspace of  $\mathcal{V}$  that is complete with the norm in (7.47), so that it is a Hilbert space.

The functional  $\mathcal{J}$  is linear by its definition and properties of integrals. It is continuous since, for  $Z_1, Z_2 \in \mathcal{V}$ , we have

$$\begin{aligned}|\mathcal{J}(Z_1) - \mathcal{J}(Z_2)| &= \left| \int_{I \times \Omega} Y(t, \omega)(Z_1(t, \omega) - Z_2(t, \omega))dt P(d\omega) \right| \\ &\leq \left( \int_{I \times \Omega} Y(t, \omega)^2 dt P(d\omega) \right)^{1/2} \left( \int_{I \times \Omega} (Z_1(t, \omega) - Z_2(t, \omega))^2 dt P(d\omega) \right)^{1/2}\end{aligned}$$

by the Cauchy–Schwarz inequality, that is,  $|\mathcal{J}(Z_1) - \mathcal{J}(Z_2)| \leq \|Y\|_{\mathcal{V}} \|Z_1 - Z_2\|_{\mathcal{V}}$ . Since  $Y \in \mathcal{V}$ , it has finite variance so that  $\mathcal{J}$  is continuous and, therefore, bounded (Theorem B.23).

The functional  $\mathcal{B}$  is bilinear by its definition. It is bounded since

$$\begin{aligned}|\mathcal{B}(X, Z)| &\leq \left| \int_{I \times \Omega} \dot{X}(t, \omega)Z(t, \omega)dt P(d\omega) \right| + \left| \int_{I \times \Omega} A(\omega)X(t, \omega)Z(t, \omega)dt P(d\omega) \right| \\ &\leq \left( \int_{I \times \Omega} \dot{X}(t, \omega)^2 dt P(d\omega) \right)^{1/2} \left( \int_{I \times \Omega} Z(t, \omega)^2 dt P(d\omega) \right)^{1/2} \\ &\quad + \left( \int_{I \times \Omega} A(\omega)^2 X(t, \omega)^2 dt P(d\omega) \right)^{1/2} \left( \int_{I \times \Omega} Z(t, \omega)^2 dt P(d\omega) \right)^{1/2} \\ &= \|\dot{X}\|_{\mathcal{V}} \|Z\|_{\mathcal{V}} + \|AX\|_{\mathcal{V}} \|Z\|_{\mathcal{V}} = \left( \|\dot{X}\|_{\mathcal{V}} + \|AX\|_{\mathcal{V}} \right) \|Z\|_{\mathcal{V}} \\ &\leq (a_2 \vee 1) \left( \|\dot{X}\|_{\mathcal{V}} + \|X\|_{\mathcal{V}} \right) \|Z\|_{\mathcal{V}} = (a_2 \vee 1) \|X\|_{\mathcal{W}} \|Z\|_{\mathcal{V}},\end{aligned}$$

by the Cauchy–Schwarz inequality and properties of random variable  $A$ .

We now show that the conditions of Theorem 7.14 are satisfied. For the first condition, take  $X \in \mathscr{W}$  and  $Z = \dot{X} + AX$ . Since  $Z \in \mathscr{V}$ , we have

$$\begin{aligned} \|Z\|_{\mathscr{V}}^2 &= \int_{I \times \Omega} (\dot{X}(t, \omega) + A(\omega)X(t, \omega))^2 dt P(d\omega) \\ &\leq 2(\|\dot{X}\|_{\mathscr{V}}^2 + \|AX\|_{\mathscr{V}}^2) = 2(a_2 \vee 1)^2 \|X\|_{\mathscr{W}}^2, \\ |\mathscr{B}(X, Z)| &= \int_{I \times \Omega} Z(t, \omega)^2 dt P(d\omega) = \|Z\|_{\mathscr{V}}^2, \text{ and} \\ \|Z\|_{\mathscr{V}}^2 &= \|\dot{X}\|_{\mathscr{V}}^2 + \|AX\|_{\mathscr{V}}^2 + 2E\left[\int_I \dot{X}(t)AX(t)dt\right] \geq (a_1 \wedge 1)^2 \|X\|_{\mathscr{W}}^2 \end{aligned}$$

by using the inequality  $(a+b)^2 \leq 2(a^2+b^2)$ ,  $a, b \in \mathbb{R}$ , the definitions of  $\mathscr{B}$  and  $Z$ , and the equality  $2E\left[\int_I \dot{X}(t)AX(t)dt\right] = E[AX(\tau)^2] \geq 0$  resulting by integration by parts with  $X(0) = 0$ . We have  $|\mathscr{B}(X, Z/\|Z\|_{\mathscr{V}})| = \|Z\|_{\mathscr{V}} \geq (\text{const})\|X\|_{\mathscr{W}}$  for  $Z \in \mathscr{V}$  of the form  $Z = \dot{X} + AX$  so that  $\sup_{Z \in \mathscr{V}, \|Z\|_{\mathscr{V}}=1} |\mathscr{B}(X, Z)| \geq (\text{const})\|X\|_{\mathscr{W}}$ ,  $\forall X \in \mathscr{W}$ , that is, the first condition in (7.45).

For the second condition in (7.45), take  $Z \in \mathscr{V}$  and set  $X(t) = \int_0^t Z(s)ds$ ,  $t \in [0, \tau]$ , so that  $\dot{X}(t) = Z(t)$  since  $\int_0^t Z(s)ds$  is defined in the m.s. sense, and  $X \in \mathscr{W}$ . We have

$$\begin{aligned} \mathscr{B}(X, Z) &= E\left[\int_0^\tau \left(Z(t) + A \int_0^t Z(s)ds\right) Z(t)dt\right] \\ &= E\left[\int_0^\tau Z(t)^2 dt\right] + E\left[A \int_0^\tau \left(\int_0^t Z(s)ds\right) Z(t)dt\right] \\ &= E\left[\int_0^\tau Z(t)^2 dt\right] + \frac{1}{2}E\left[A\left(\int_0^\tau Z(t)dt\right)^2\right] \geq 0 \end{aligned}$$

by using integration by parts,  $E\left[\left(\int_0^\tau Z(t)dt\right)^2\right] \geq 0$ , and  $E\left[A \int_0^\tau Z(t)^2 dt\right] \geq 0$ . The second condition in (7.45) follows from the observation that  $|\mathscr{B}(X, Z)| > 0$  for  $Z \in \mathscr{V} \setminus \{0\}$  and that supremum of  $|\mathscr{B}(X, Z)|$  taken over all  $X \in \mathscr{W}$  is larger than  $|\mathscr{B}(X, Z)|$  for the definition of  $X$  used in these calculations.  $\blacktriangle$

*Example 7.25* Suppose that  $A$  in Example 7.24 is a real-valued stochastic process  $A(t)$ , so that  $X(t)$  is the solution of  $\dot{X}(t) + A(t)X(t) = Y(t)$ ,  $t \in I = [0, \tau]$ , with  $X(0) = 0$ . It is assumed that  $A(t)$  and  $Y(t)$  are independent processes defined on a probability space  $(\Omega, \mathscr{F}, P)$  and that  $P(\inf_{t \in I} A(t) \geq a_1, \sup_{t \in I} A(t) \leq a_2) = 1$  for some constants  $0 < a_1 \leq a_2 < \infty$ . Let  $\mathscr{W}$ ,  $\mathscr{V}$ , and  $\mathscr{S}$  be as in Example 7.24, and let  $\mathscr{B} : \mathscr{W} \times \mathscr{V} \rightarrow \mathbb{R}$  be given by

$$\mathscr{B}(X, Z) = E\left[\int_I \left(\dot{X}(t) + A(t)X(t)\right) Z(t)dt\right], \quad (7.49)$$

which constitute a direct extension of  $\mathscr{B}$  in (7.48).

We have seen in Example 7.24 that  $\mathcal{J}$  is a linear functional that is continuous and, therefore, bounded. That  $\mathcal{B}$  is bilinear follows from its definition. Arguments similar to those used in Example 7.24 show that  $\mathcal{B}$  is bounded and weakly coercive, so that the stochastic equation  $\dot{X}(t) + A(t)X(t) = Y(t)$ ,  $t \in I = [0, \tau]$ , admits a unique weak solution (Theorem 7.14).  $\diamond$

*Proof* The set  $\mathcal{W}$  is included in  $\mathcal{V}$  since, for  $X \in \mathcal{W}$  and almost all  $\omega \in \Omega$ , we have

$$\begin{aligned} |X(t, \omega)| &= \left| \int_0^t \dot{X}(s, \omega) ds \right| \leq \left( \int_0^t ds \right)^{1/2} \left( \int_0^t \dot{X}(s, \omega)^2 ds \right)^{1/2} \\ &\leq \tau^{1/2} \left( \int_0^\tau \dot{X}(s, \omega)^2 ds \right)^{1/2} = \tau^{1/2} \|\dot{X}(\cdot, \omega)\|_{L^2(I)}, \end{aligned}$$

by the Cauchy–Schwarz inequality, or  $\|X(\cdot, \omega)\|_{L^2(I)} \leq \tau \|\dot{X}(\cdot, \omega)\|_{L^2(I)}$  a.s., which gives  $\|X\|_{\mathcal{V}} \leq \tau \|\dot{X}\|_{\mathcal{V}}$  by expectation. Hence,  $X \in \mathcal{W}$  implies  $\|X\|_{\mathcal{V}} < \infty$ , that is,  $X \in \mathcal{V}$ .

For  $X \in \mathcal{W}$  and  $Z \in \mathcal{V}$ , we have

$$\begin{aligned} |\mathcal{B}(X, Z)| &= \left| \int_{I \times \Omega} \left( \dot{X}(t, \omega) Z(t, \omega) + A(t, \omega) X(t, \omega) Z(t, \omega) \right) dt P(d\omega) \right| \\ &\leq \|\dot{X}\|_{\mathcal{V}} \|Z\|_{\mathcal{V}} + \left| \int_{I \times \Omega} A(t, \omega) X(t, \omega) Z(t, \omega) dt P(d\omega) \right| \\ &\leq \|\dot{X}\|_{\mathcal{V}} \|Z\|_{\mathcal{V}} + \|AX\|_{\mathcal{V}} \|Z\|_{\mathcal{V}} = (\|\dot{X}\|_{\mathcal{V}} + \|AX\|_{\mathcal{V}}) \|Z\|_{\mathcal{V}} \\ &\leq (a_2 \vee 1) \|X\|_{\mathcal{W}} \|Z\|_{\mathcal{V}}, \end{aligned}$$

that is,  $\mathcal{B}$  is bounded.

For the first condition in (7.45), take  $X \in \mathcal{W}$  and set  $Z = \dot{X} + AX \in \mathcal{V}$ . Direct calculations give

$$\begin{aligned} \|Z\|_{\mathcal{V}}^2 &= \int_{I \times \Omega} \left( \dot{X}(t, \omega) + A(t, \omega) X(t, \omega) \right)^2 dt P(d\omega) \\ &\leq 2(\|\dot{X}\|_{\mathcal{V}}^2 + \|AX\|_{\mathcal{V}}^2) \leq 2(a_2 \vee 1)^2 \|X\|_{\mathcal{W}}^2, \\ |\mathcal{B}(X, Z)| &= E \left[ \int_I (\dot{X}(t) + A(t)X(t))^2 dt \right] = \|Z\|_{\mathcal{V}}^2, \quad \text{and} \\ \|Z\|_{\mathcal{V}}^2 &= \|\dot{X}\|_{\mathcal{V}}^2 + \|AX\|_{\mathcal{V}}^2 + 2E \left[ \int_I A(t)X(t)\dot{X}(t) dt \right]. \end{aligned}$$

The expectation  $E \left[ \int_I A(t)X(t)\dot{X}(t) dt \right]$  is bounded since  $\|\dot{X}\|_{\mathcal{V}}^2 + \|AX\|_{\mathcal{V}}^2 \geq 0$  and  $\|Z\|_{\mathcal{V}}^2 \geq 0$  are finite. Let  $c > 0$  be a constant such that  $\|Z\|_{\mathcal{V}}^2 \geq c^2 \|X\|_{\mathcal{W}}^2$ , then  $|\mathcal{B}(X, Z/\|Z\|_{\mathcal{V}})| = \|Z\|_{\mathcal{V}} \geq c \|X\|_{\mathcal{W}}$  for  $Z = \dot{X} + AX$ , as in the previous example. This implies  $\sup_{Z \in \mathcal{V}, \|Z\|_{\mathcal{V}}=1} |\mathcal{B}(X, Z)| \geq (\text{const}) \|X\|_{\mathcal{W}}$  for all  $X \in \mathcal{W}$ .

For the second condition in (7.45), set  $X(t) = \int_0^t A(s)Z(s)ds$  for an arbitrary  $Z \in \mathcal{V}$ . Then

$$\begin{aligned}\mathcal{B}(X, Z) &= E \left[ \int_I \left( \dot{X}(t) + A(t) \int_0^t A(s)Z(s)ds \right) Z(t)dt \right] \\ &= E \left[ \int_I A(t)Z(t)^2 dt \right] + E \left[ \int_I A(t)Z(t) \left( \int_0^t A(s)Z(s)ds \right) dt \right] \\ &= E \left[ \int_I A(t)Z(t)^2 dt \right] + \frac{1}{2} E \left[ \left( \int_I A(t)Z(t)dt \right)^2 \right],\end{aligned}$$

by performing integration by parts on the integral  $\int_I A(t)Z(t) \left( \int_0^t A(s)Z(s)ds \right) dt$ . Since  $A(t) > 0$  a.s., we have  $\mathcal{B}(X, Z) \geq 0$  for arbitrary  $Z \in \mathcal{V}$  and  $\mathcal{B}(X, Z) > 0$  for  $Z \in \mathcal{V} \setminus \{0\}$ . Since  $Z$  is arbitrary, the second condition in (7.45) holds.  $\blacktriangle$

Recall that we have established in Sect. 5.5.1 conditions for the existence and uniqueness of strong solutions for stochastic differential equations driven by Gaussian and semimartingale noise. The theorems in Sect. 5.5.1 offer alternatives to Theorem 7.14 since they can also be applied to differential equations with uncertain coefficients and colored driving noise. For example, if  $Y(t) \stackrel{!}{=} dB(t)/dt$  in Example 7.24 is a Gaussian white noise, then  $\dot{X}(t) + AX(t) = Y(t)$  in this example can be given in the form

$$\begin{cases} dX_1(t) = -X_2(t) X_1(t) dt + dB(t) \\ dX_2(t) = 0, \end{cases}$$

with initial conditions  $X_1(0) = 0$  and  $X_2(0) = A$ , where  $(X_1 = X, X_2 = A)$ . The diffusion process  $(X_1, X_2)$  defined by this equations is unique in the strong sense if its drift and diffusion coefficients satisfy the conditions in Theorem 5.8. If  $Y(t)$  in Example 7.24 is a colored noise that can be described by the output of a linear filter driven by white noise, then  $(X_1, X_2)$  augmented with the state vector of this filter satisfies a stochastic differential equation driven by white noise.

We conclude this section with an outline of the methods for solving approximately differential equations with random coefficients and input discussed in the following sections. The Monte Carlo, conditional analysis, state augmentation, stochastic reduced order models, stochastic Galerkin, and stochastic collocation methods are applied in Sects. 7.4.2–7.4.8 to solve stochastic differential equations with random coefficients of arbitrary uncertainty. The methods in Sect. 7.4.9 are for stochastic differential equations with random coefficients of small uncertainty.

## 7.4.2 Monte Carlo Simulation

Monte Carlo simulation is the most general method for solving SDEs with random coefficients. Its implementation requires complete information on the probability law

of both random input and coefficients. Deterministic solvers can be used to calculate samples of the solutions of SDEs from samples of their random coefficients and input processes. The resulting samples can be used to estimate moments and other properties of the solutions of these equations.

Note that Monte Carlo simulation delivers samples of strong solutions of SDEs since they correspond to input/coefficient samples and the generation of these samples requires to select versions for all random elements. For example, the generation of samples of the real-valued process  $X(t)$  defined by  $dX(t) = -AX(t)dt + dB(t)$  in Example 7.23 requires to specify the version of the Brownian motion  $B(t)$  and the probability law of  $A$ . Also, the calculation of solution samples requires that the random elements of SDEs be such that these equations have unique solutions in the strong sense (Sect. 5.5.1).

### 7.4.3 Conditional Analysis

Let  $X(t)$  be the solution of (7.2) and assume that the probability law of conditional process  $X(t) \mid \Theta$  is available analytically or constitutes the output of an efficient algorithm.

*Example 7.26* Let  $X(t)$  be as in Example 7.23, so that  $\Theta = A$  and  $X(t) \mid A$  is an Ornstein–Uhlenbeck process, that is, a Gaussian process with mean  $\mu(t; A) = \mu_0 \exp(-At)$ , variance function  $\gamma(t; A) = \gamma_0 \exp(-2At) + (1 - \exp(-2At))/(2A)$ , and covariance function  $c(s, t; A) = \gamma(s \wedge t; A) \exp(-A|s - t|)$ . The density of the conditional vector  $(X(t_1), \dots, X(t_n)) \mid A$  is

$$f(x; A) = [(2\pi)^n \det(\gamma(A))]^{-1/2} \exp \left[ -\frac{1}{2} (x - \mu(A))' \gamma(A)^{-1} (x - \mu(A)) \right]$$

where  $x = (x_1, \dots, x_n)'$ ,  $\mu(A) = (\mu(t_1; A), \dots, \mu(t_n; A))'$ , and  $\gamma(A) = \{c(t_i, t_j; A), i, j = 1, \dots, n\}$ .

Properties of the unconditional solution  $X(t)$  can be obtained from the second moment properties and the finite dimensional densities of  $X(t) \mid A$  by direct integration or Monte Carlo simulation. For example, the mean of  $X(t)$  is the expectation  $\mu(t) = E[\mu_0 \exp(-At)]$ , that can be obtained by numerical integration or can be estimated from

$$\hat{\mu}(t) = \frac{1}{n_s} \sum_{k=1}^{n_s} \mu_0 \exp(-a_k t)$$

where  $\{a_k, k = 1, \dots, n_s\}$  are  $n_s$  independent samples of  $A$ .  $\diamond$

The solution in Example 7.26 by conditional analysis is efficient since the probability law of  $X(t) \mid A$  is known as a function of  $\Theta = A$ . Generally, the probability law of  $X(t) \mid \Theta$  is not available analytically, so that this approach is rarely useful in

applications. For example, the probability law of  $X(t) \mid \Theta$  in Example 7.26 cannot be expressed as a function of  $\Theta = A$  if the Gaussian driving noise is replaced with an arbitrary non-Gaussian process.

#### 7.4.4 Conditional Monte Carlo Simulation

Consider a special case of (7.1) with real-valued state  $X(t)$  and coefficients that do not depend explicitly on time. We denote  $Y(t)$  in (7.1) by  $\xi(t)$  since this process has some special features. Specifically,  $X(t)$  is the solution of

$$dX(t) = a(X(t), \xi(t)) dt + b(X(t), \xi(t)) dB(t), \quad t \geq 0, \quad (7.50)$$

where  $B(t)$  denotes a standard Brownian motion and  $\xi(t)$  is a semi-Markov process with values in a finite set  $\{\theta_1, \dots, \theta_n\}$  and transition rates  $\{q_{lk}(t), k, l = 1, \dots, n\}$  from state  $\theta_l$  to state  $\theta_k$  at time  $t$ . Let  $T_0 < T_1 < \dots < T_r < \dots$  denote random transition times defined by the recurrence formula

$$T_r = T_{r-1} + S_r, \quad r = 1, 2, \dots, \quad (7.51)$$

where  $T_0 = 0$  and  $\{S_r\}$  are iid random variables with finite mean. Since  $\xi(t)$  is constant in  $[T_{r-1}, T_r)$ ,  $r = 1, 2, \dots$ ,  $X(t)$  in (7.50) is a diffusion process during the time intervals  $[T_{r-1}, T_r)$  with drift and diffusion coefficients  $a(X(t), \xi(t))$  and  $b(X(t), \xi(t))$ . It is assumed that the drift and diffusion coefficients are such that (7.50) has a solution that is unique for all values of  $\xi(t)$  (Sect. 5.5.1.1 in this book, [4], Sect. 4.7.1.1).

We present two methods for calculating properties of  $X(t)$ . The first method proposed in [7–11] develops integral equations for marginal moments and other properties of  $X(t)$  under some rather restrictive conditions. The second method, referred to as conditional Monte Carlo simulation, uses the random vibration theory to find properties of the conditional process  $X(t) \mid \xi(t)$  in each time interval  $[T_{r-1}, T_r)$ . Unconditional properties of  $X(t)$  are obtained by averaging properties of  $X(t) \mid \xi(t)$  over samples of  $\xi(t)$ . Conditional Monte Carlo simulation is particularly efficient for stochastic differential equations with drift and diffusion coefficients that are linear in and independent of  $X(t)$ , respectively. In this case,  $X(t) \mid \xi(\cdot)$  is Gaussian in the intervals of constant values of  $\xi(t)$ .

We now review the first method that delivers integral equations for moments and other state properties. Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function and denote by

$$G_k(t, x) = E[h(X(t)) \mid X(0) = x, \xi(0) = \theta_k], \quad t \geq 0, \quad (7.52)$$

the expectation of  $h(X(t))$  conditional on  $(X(0) = x, \xi(0) = \theta_k)$  calculated on a sample of Brownian motion  $B(t)$ . The latter condition is not indicated explicitly for simplicity. Depending on the functional form of  $h$ ,  $G_k(t, x)$  provides various properties for  $X(t)$ . For example,  $G_k(t, x)$  is the conditional expectation of  $X(t)$  if



$h(x) = x$ , the real or the imaginary part of the conditional characteristic function of  $X(t)$  if  $h(x) = \cos(ux)$  or  $h(x) = \sin(ux)$ ,  $u \in \mathbb{R}$ , and the conditional marginal distribution of  $X(t)$  at  $z$  if  $h(x) = 1(x \leq z)$ .

We follow the arguments in [7–11] to develop an integral equation for  $E[X(t) | X(0) = x, \xi(0) = \theta_k]$ . If  $\xi$  has no transition in  $[0, t)$  and  $\xi(0) = \theta_k$ , then  $X(t) = X_k(t)$  satisfies the stochastic differential equation  $dX_k(t) = a_k(X_k(t))dt + b_k(X_k(t))dB(t)$ ,  $t \geq 0$ , where  $a_k(\cdot) = a(\cdot, \theta_k)$  and  $b_k(\cdot) = b(\cdot, \theta_k)$ . The probability of this event is  $1 - \int_0^t Q_k(\tau)d\tau$ , where

$$Q_k(\tau) = \sum_{l=1, l \neq k}^n q_{kl}(\tau), \quad (7.53)$$

is the rate of transition out of  $\theta_k$  at time  $\tau$ .

Suppose  $\xi$  has a jump at time  $\tau \in [0, t)$ , and its value changes from  $\theta_k$  to  $\theta_l$ . The rate of this change at time  $\tau$  is  $q_{kl}(\tau)$ . The expectation of  $h(X(t))$  conditional on  $(X(\tau), \xi(\tau) = \theta_l)$  is  $G_l(t - \tau, X(\tau))$ ,  $t \geq \tau$ , on a sample of Brownian motion  $B(t)$ , so that

$$\begin{aligned} G_k(t, x) &= E \left[ h(X(t)) \mid X(0) = x, \xi(0) = \theta_k, \xi \text{ has no jump in } [0, t) \right] \left( 1 - \int_0^t Q_k(\tau)d\tau \right) \\ &\quad + \int_0^t \sum_{l=1, l \neq k}^n E \left[ h(X(t)) \mid X(0) = x, \xi(0) = \theta_k, \xi \text{ jumps to } \theta_l \text{ at } \tau \in [0, t) \right] q_{kl}(\tau)d\tau \\ &= h(X_k(t)) \left( 1 - \int_0^t Q_k(\tau)d\tau \right) + \int_0^t \sum_{l=1, l \neq k}^n G_l(t - \tau, X_k(\tau)) q_{kl}(\tau)d\tau \end{aligned} \quad (7.54)$$

by arguments of the renewal theory ([29], Chap. 3).

Generally, (7.54) cannot be solved analytically. Numerical solutions of this equation are impractical since they require discretization of both spatial and temporal coordinates,  $X_k(\tau)$  can be a  $d$ -dimensional process,  $d \geq 1$ , and  $G_k(t, x)$  corresponds to a single sample of the driving noise. It seems that (7.54) can be solved efficiently only if the drift  $a(X(t), \xi(t))$  is linear in  $X(t)$  and  $b(X(t), \xi(t)) = 0$ , that is, a linear dynamic system with no driving noise. In this case, we have  $G_k(t, x) = xG_k(t)$  and  $G_l(t - \tau, X_k(\tau)) = X_k(\tau)G_l(t - \tau)$  so that (7.54) becomes

$$xG_k(t) = h(X_k(t)) \left( 1 - \int_0^t Q_k(\tau)d\tau \right) + \int_0^t \sum_{l=1, l \neq k}^n X_k(\tau) G_l(t - \tau) q_{kl}(\tau)d\tau. \quad (7.55)$$

The expectations  $G_k(t)$  can be obtained numerically from, for example, the Laplace transform of (7.55), as illustrated in [9].

Consider now the second method, refer to as conditional Monte Carlo simulation. The method calculates properties of the state of differential and difference equations with random coefficients and input in two steps. First, random vibration theory is

used to calculate properties of the conditional process  $X(t) \mid \xi(\cdot)$  on samples of  $\xi(t)$ . Second, unconditional properties of  $X(t)$  are obtained by averaging conditional properties of  $X(t) \mid \xi(\cdot)$  over samples of  $\xi(t)$ .

Let  $\xi(t, \omega)$  be a sample of the semi-Markov process  $\xi(t)$  in (7.50) defined by a sequence of jump times  $0 = T_0(\omega) < T_1(\omega) < \dots < T_r(\omega) < \dots$  and states  $\xi_r(\omega) = \xi(T_r(\omega), \omega) \in \{\theta_1, \dots, \theta_n\}$ . Since  $\xi(t, \omega)$  is constant during the time intervals  $[T_{r-1}(\omega), T_r(\omega))$ ,  $r = 1, 2, \dots$ , we have

$$\xi(t, \omega) = \sum_{r \geq 1} \xi_r(\omega) 1(T_{r-1}(\omega) \leq t < T_r(\omega)), \quad t \geq 0. \quad (7.56)$$

We have seen that the solution  $X_r(t)$  of (7.50) with  $\xi(t, \omega)$  in place of  $\xi(t)$  in  $[T_{r-1}(\omega), T_r(\omega))$  is a diffusion process defined by the stochastic differential equation

$$dX_r(t) = a_r(X_r(t)) dt + b_r(X_r(t)) dB(t), \quad t \in [T_{r-1}(\omega), T_r(\omega)), \quad (7.57)$$

with initial state  $X_{r-1}(T_{r-1}(\omega))$ , drift  $a_r(\cdot) = a(\cdot, \xi_r(\omega))$ , and diffusion  $b_r(\cdot) = b(\cdot, \xi_r(\omega))$ . If the drift and diffusion coefficients in (7.50) are such that this equation has a unique solution almost surely, the processes satisfying (7.57) exist and are uniquely defined. The notation  $X_r(t)$  may be misleading since it does not emphasize the dependence on  $\xi(t, \omega)$ , that is, the fact that  $X_r(t)$  is the conditional process  $X(t) \mid \xi(\cdot, \omega)$  restricted to time interval  $[T_{r-1}(\omega), T_r(\omega))$ . We use this notation for simplicity.

Efficient methods for calculating statistics of the conditional processes  $\{X_r(t), r = 1, 2, \dots\}$  are available for arbitrary equations. However, properties of these processes can be obtained efficiently in cases of practical interest, for example, differential equations with linear drift and state-invariant diffusion, that is equations of the type

$$dX_r(t) = a_r X_r(t) dt + b_r dB(t), \quad t \in [T_{r-1}(\omega), T_r(\omega)), \quad (7.58)$$

where  $a_r$  and  $b_r$  denote  $a(\xi(t))$  and  $b(\xi(t))$  in the time interval  $[T_{r-1}(\omega), T_r(\omega))$ . Equations of this type are studied in linear random vibration. The probability law of the Gaussian processes  $X_r(t)$  is completely defined by their second moment properties, that can be obtained from the mean and covariance equations in (7.6).

We first illustrate the second method by the solution  $X(t)$  of the stochastic differential equation

$$dX(t) = -\xi(t) X(t) dt + dB(t), \quad t \geq 0, \quad (7.59)$$

that is, a linear version of (7.50) with drift  $a(X(t), \xi(t)) = -\xi(t)X(t)$ ,  $\xi(t) > 0$ , diffusion  $b(X(t), \xi(t)) = 1$ , and initial state  $X(0) = X_0$ , a Gaussian variable with mean  $\mu_0$  and variance  $\gamma_0$  that is independent of  $B(t)$ .

**Theorem 7.15** *Let  $X(t)$  be defined by (7.59). The mean and the variance functions,  $\mu_r(t)$  and  $\gamma_r(t)$ , of the conditional processes  $\{X_r(t)\}$  on a sample  $\xi(t, \omega)$  of  $\xi(t)$  are*

$$\begin{aligned}\mu_r(t) &= \mu_{r-1}(T_{r-1}(\omega))e^{-\xi_r(\omega)(t-T_{r-1}(\omega))} \\ \gamma_r(t) &= \gamma_{r-1}(T_{r-1}(\omega))e^{-2\xi_r(\omega)(t-T_{r-1}(\omega))}\end{aligned}\quad (7.60)$$

for  $t \in [T_{r-1}(\omega), T_r(\omega)]$ , so that the recurrence formulas

$$\begin{aligned}\mu_r(T_r(\omega)) &= \mu_{r-1}(T_{r-1}(\omega))e^{-\xi_r(\omega)(T_r(\omega)-T_{r-1}(\omega))} \\ \gamma_r(T_r(\omega)) &= \gamma_{r-1}(T_{r-1}(\omega))e^{-2\xi_r(\omega)(T_r(\omega)-T_{r-1}(\omega))}\end{aligned}\quad (7.61)$$

hold for  $r = 1, 2, \dots$  with  $\mu_0(T_0(\omega)) = \mu_0$  and  $\gamma_0(T_0(\omega)) = \gamma_0$ .

*Proof* Results in (7.60) are solutions of (7.6). For example, the mean equation is  $\dot{\mu}_r(s) = -\xi_r(\omega)\mu_r(s)$ ,  $s \in [0, T_r(\omega) - T_{r-1}(\omega)]$ , with initial condition  $\mu_r(0) = \mu_{r-1}(T_{r-1}(\omega))$ .  $\blacktriangle$

**Theorem 7.16** Let  $X(t)$  be defined by (7.59). The correlations of the random variables  $X_q(T_q(\omega))$  and  $X_p(T_p(\omega))$ ,  $0 \leq p < q$ , can be obtained recursively from

$$E[X_q(T_q(\omega))X_p(T_p(\omega))] = E[X_{q-1}(T_{q-1}(\omega))X_p(T_p(\omega))]e^{-\xi_q(\omega)(T_q(\omega)-T_{q-1}(\omega))}. \quad (7.62)$$

*Proof* Note that

$$X_q(T_q(\omega)) = X_{q-1}(T_{q-1}(\omega))e^{-\xi_q(\omega)(T_q(\omega)-T_{q-1}(\omega))} + \int_{T_{q-1}(\omega)}^{T_q(\omega)} e^{-\xi_q(\omega)(T_q(\omega)-s)} dB(s),$$

so that the expectation of the product  $X_q(T_q(\omega))X_p(T_p(\omega))$  has the expression in (7.62) since  $X_p(T_p(\omega))$  is independent of future values of the Brownian motion. The above formula can be applied recursively beginning with  $q = p + 1$  since  $E[X_q(T_q(\omega))X_p(T_p(\omega))]$  for this value of  $q$  depends on  $E[X_p(T_p(\omega))X_p(T_p(\omega))]$ , which is given by (7.61).  $\blacktriangle$

**Theorem 7.17** Let  $X(t)$  be defined by (7.59). The correlation function of the conditional processes  $\{X_r(t)\}$  is given by

$$\begin{aligned}E[X_p(s)X_q(t)] &= E[X_{p-1}(T_{p-1}(\omega))X_{q-1}(T_{q-1}(\omega))]e^{-\xi_p(\omega)(s-T_{p-1}(\omega))}e^{-\xi_q(\omega)(t-T_{q-1}(\omega))} \\ &\quad + E\left[X_{q-1}(T_{q-1}(\omega))\int_{T_{p-1}(\omega)}^s e^{-\xi_p(\omega)(s-u)} dB(u)\right]e^{-\xi_q(\omega)(t-T_{q-1}(\omega))}\end{aligned}\quad (7.63)$$

for  $0 \leq p < q$ ,  $s \in [T_{p-1}(\omega), T_p(\omega)]$ ,  $t \in [T_{q-1}(\omega), T_q(\omega)]$ .

*Proof* Straightforward calculations yield (7.63). For example,  $E[X_p(s)X_q(t)]$  for  $q = p + 1$  is

$$\begin{aligned}
& E[X_p(s)X_{p+1}(t)] \\
&= E\left[\left(X_{p-1}(T_{p-1}(\omega))e^{-\xi_p(\omega)(s-T_{p-1}(\omega))} + \int_{T_{p-1}(\omega)}^s e^{-\xi_p(\omega)(s-u)} dB(u)\right)\right. \\
&\quad \times \left.\left(X_p(T_p(\omega))e^{-\xi_{p+1}(\omega)(t-T_p(\omega))} + \int_{T_p(\omega)}^t e^{-\xi_{p+1}(\omega)(t-v)} dB(v)\right)\right] \\
&= E\left[X_{p-1}(T_{p-1}(\omega))X_{p-1}(T_p(\omega))\right]e^{-\xi_p(\omega)(s-T_{p-1}(\omega))}e^{-\xi_{p+1}(\omega)(t-T_p(\omega))} \\
&\quad + E\left[X_p(T_p(\omega))\int_{T_{p-1}(\omega)}^s e^{-\xi_p(\omega)(s-u)} dB(u)\right]e^{-\xi_{p+1}(\omega)(t-T_p(\omega))}
\end{aligned}$$

and

$$\begin{aligned}
& E\left[X_p(T_p(\omega))\int_{T_{p-1}(\omega)}^s e^{-\xi_p(\omega)(s-u)} dB(u)\right] \\
&= E\left[\int_{T_{p-1}(\omega)}^{T_p(\omega)} e^{-\xi_p(\omega)(T_p(\omega)-v)} dB(v)\int_{T_{p-1}(\omega)}^s e^{-\xi_p(\omega)(s-u)} dB(u)\right] \\
&= \int_{T_{p-1}(\omega)}^s e^{-\xi_p(\omega)(T_p(\omega)-u)} e^{-\xi_p(\omega)(s-u)} du = e^{-\xi_p(\omega)(T_p(\omega)+s)} \int_{T_{p-1}(\omega)}^s e^{-2\xi_p(\omega)u} du \\
&= e^{-\xi_p(\omega)(T_p(\omega)+s)} \left(e^{2\xi_p(\omega)s} - e^{2\xi_p(\omega)T_{p-1}(\omega)}\right) / (2\xi_p(\omega)).
\end{aligned}$$

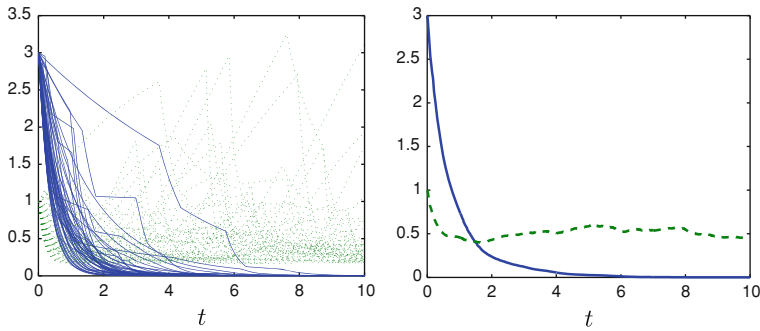
Similar calculations can be performed to find correlations  $E[X_p(s)X_q(t)]$  for arbitrary times  $s$  and  $t$ . ▲

We conclude the analysis of the stochastic differential equation in (7.59) with the following comments. First, Theorems 7.15–7.17 hold for arbitrary non-Gaussian noise equal in the second moment sense with the Brownian motion  $B(t)$ , for example, Poisson white noise defined as the formal derivative of a compound Poisson process  $C(t) = \sum_{k=1}^{N(t)} Y_k$ , where  $N(t)$  is a homogeneous Poisson process with intensity  $\rho > 0$  and  $\{Y_k\}$  denote iid real-valued random variables with mean  $E[Y_1] = 0$  and finite variance such that  $\rho E[Y_1^2] = 1$ . Second, these theorems can be extended to  $\mathbb{R}^d$ -valued state vectors  $X(t)$ . Third, similar arguments can be used to calculate statistics for solutions  $X(t)$  of nonlinear stochastic equations. The calculations are feasible if properties of  $X(t)$  conditional on  $\xi(t)$  can be obtained with a reasonable effort. For example, let  $X(t)$  be a geometric Brownian motion defined by the stochastic differential equation

$$dX(s) = c X(s) dt + \sigma X(s) dB(s), \quad s \in [0, T_r(\omega) - T_{r-1}(\omega)), \quad (7.64)$$

where  $c$  and  $\sigma$  correspond to a sample  $\xi(t, \omega)$  of  $\xi(t)$  during a time interval  $[T_{r-1}(\omega), T_r(\omega))$ . The solution of (7.64) is (Example 5.6)

$$X(s) = X(0)e^{(c-\sigma^2/2)s+\sigma B(s)}, \quad s \in [0, T_r(\omega) - T_{r-1}(\omega)), \quad (7.65)$$



**Fig. 7.11** Conditional means and variances of  $X(t)$  (left panel) and mean and variance of  $X(t)$  (right panel) for  $(a = 0.01, b = 3.0)$

where the initial state  $X(0) = X(T_{r-1}(\omega))$  is a random variable independent of  $\{B(s), s \geq 0\}$ . Statistics of  $X(t)$  conditional on  $\xi(t)$  can be obtained simply. For example, the marginal distribution of  $X(s)$  in  $[0, T_r(\omega) - T_{r-1}(\omega))$  conditional on  $X(T_{r-1}(\omega)) = x_0$  is

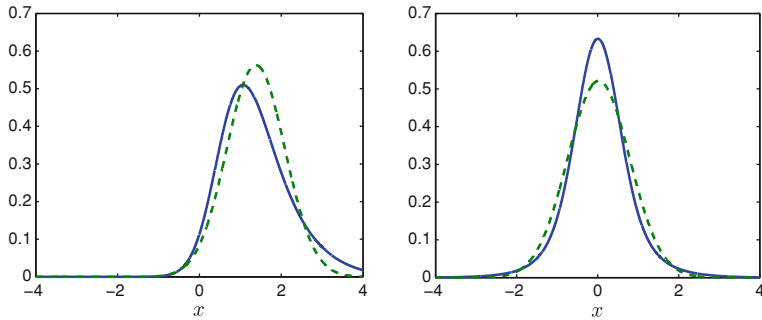
$$F(x | x_0) = \Phi\left(\frac{(\ln(x/x_0) - (c - \sigma^2/2)s)/(\sigma\sqrt{s})}{1}\right), \quad s \in [0, T_r(\omega) - T_{r-1}(\omega)). \quad (7.66)$$

The unconditional marginal distribution results by averaging over  $X(0)$ . Other examples of nonlinear stochastic differential equations that admit simple solutions can be found in [41] (Sect. 4.4).

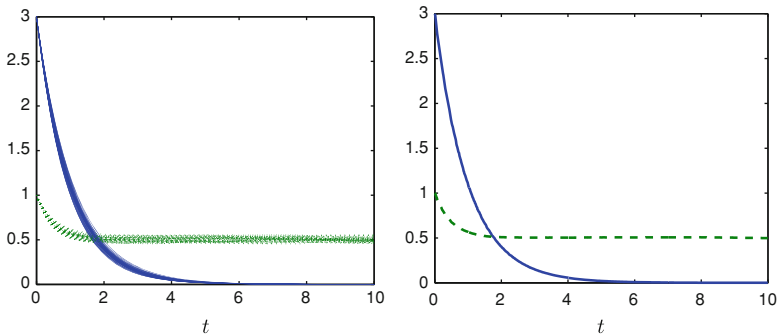
The remainder of this section presents two examples showing that the second method works even if  $\xi(t)$  does not take a finite number of values and can be used to assess the performance of degrading systems.

**Example 7.27** Let  $X(t)$  be defined by (7.59) with initial state  $X(0) \sim N(\mu_0, \gamma_0)$  assumed to be independent of  $B(t)$  and  $\xi(t)$ . The semi-Markov process  $\xi(t)$  takes independent  $U(a, b)$  values in distinct time intervals  $[T_{r-1}, T_r)$ , where  $T_r$  are the jump times of a homogeneous Poisson process with intensity  $\lambda > 0$ . Numerical results are for  $\mu_0 = 3$ ,  $\sigma_0 = 1$ , and  $\lambda = 1$ . The plots in Figs. 7.11–7.12 and Fig. 7.13 are for  $(a = 0.01, b = 3.0)$  and  $(a = 0.9, b = 1.1)$ , respectively, and are based on  $n_s = 50$  independent samples of  $\xi(t)$ .

Conditional means and variances of  $X(t)$  are shown with solid and dotted lines in the left panel of Fig. 7.11. The solid and dotted lines in the right panel of the figure are averages of the conditional means and variances, that is, estimates of the mean and variance of  $X(t)$ . Figure 7.12 shows with solid lines estimates of the marginal density of  $X(t)$  at times  $t = 0.5$  (left panel) and  $t = 5$  (right panel). The dotted lines in the figure are Gaussian densities matching the means and variances of  $X(t)$  at these times. They show that  $X(t)$  is not a Gaussian process and that the discrepancy between the marginal distribution of  $X(t)$  and the Gaussian distribution changes in time.



**Fig. 7.12** Marginal densities of  $X(t)$  at times  $t = 0.5$  (left panel) and  $t = 5$  (right panel) for  $(a = 0.01, b = 3.0)$



**Fig. 7.13** Conditional means and variances of  $X(t)$  (left panel) and mean and variance of  $X(t)$  (right panel) for  $(a = 0.9, b = 1.1)$

Figure 7.13 shows plots similar to those in Fig. 7.11 for  $\xi(t)$  taking values in the interval  $[0.9, 1.1]$  rather than  $[0.01, 3.0]$ . In contrast to Fig. 7.11, differences between conditional means and variances calculated from  $n_s = 50$  independent samples of  $\xi(t)$  are small. Marginal distributions of  $X(t)$  are indistinguishable from Gaussian distributions with the means and variances of  $X(t)$ , and are not shown. This is an expected result since  $X(t)$  becomes an Ornstein–Uhlenbeck process as  $|b - a| \rightarrow 0$ .

If  $\lambda \rightarrow 0$ , the semi-Markov process  $\xi(t)$  is the random variable  $\xi(t) = \xi(0)1(t \geq 0)$  so that the marginal density of  $X(t)$  is

$$f_{X(t)}(x) = \int_a^b \frac{1}{\sigma_{X(t)}(\alpha)} \phi\left(\frac{x - \mu_{X(t)}(\alpha)}{\sigma_{X(t)}(\alpha)}\right) \frac{d\alpha}{b - a} \quad (7.67)$$

where  $\mu_{X(t)}(\alpha)$  and  $\sigma_{X(t)}(\alpha)$  denote the mean and standard deviation of  $X(t)$  for  $\xi(0) = \alpha$ . Generally,  $f_{X(t)}(x)$  is not a Gaussian density.  $\diamond$

**Example 7.28** Consider a linear oscillator with mass  $M$ , damping  $C$ , and stiffness  $K$  subjected to Gaussian white noise with mean 0 and one-sided spectral density of intensity  $g_0 > 0$ . The oscillator displacement satisfies the differential equation

$$M\ddot{X}(t) + C\dot{X}(t) + KX(t) = W(t), \quad t \geq 0, \quad (7.68)$$

where  $W(t)$  is interpreted as the formal derivative of standard Brownian motion  $B(t)$  scaled by  $\sqrt{\pi g_0}$ , that is,  $W(t) \stackrel{!}{=} \sqrt{\pi g_0} dB(t)/dt$ .

It is assumed that (1)  $K$  can be in one of the states  $k_1 > \dots > k_r \dots > k_{m+1} > 0$ , where  $m \geq 1$  is an integer and  $X^{(r)}(t)$ , defined by

$$M\ddot{X}^{(r)}(t) + C\dot{X}^{(r)}(t) + k_r X^{(r)}(t) = W(t), \quad r = 1, \dots, m, \quad (7.69)$$

denotes the oscillator displacement in damage state  $k_r$ , (2) transitions of  $K$  are only possible from  $k_r$  to  $k_{r+1}$ , and they occur when the system state leaves a safe set  $(-x_{\text{cr}}, x_{\text{cr}})$ ,  $x_{\text{cr}} > 0$ , and (3)  $x_{\text{cr}}$  is sufficiently high such that the random times  $\Gamma_r$  of residence in damage state  $k_r$  are much longer than the duration of transient system response in this state, so that the crossings of  $X^{(r)}(t)$  out of  $(-x_{\text{cr}}, x_{\text{cr}})$  are rare events. Accordingly, these crossings define approximately a homogeneous Poisson process with intensity

$$\mu_r(x_{\text{cr}}) = \frac{v_r}{\pi} \exp\left(-\frac{x_{\text{cr}}^2}{2\sigma_r^2}\right) \quad (7.70)$$

for  $M = 1$ , where  $\sigma_r^2 = \pi g_0 / (4\zeta_r v_r^3)$  denotes the stationary variance of  $X^{(r)}(t)$ ,  $v_r^2 = k_r$ , and  $2\zeta_r v_r = C$  ([42], Sect. 12.2), and the distribution of  $\Gamma_r$  can be calculated from

$$P(\Gamma_r > s) \simeq \exp(-\mu_r(x_{\text{cr}})s), \quad s \geq 0. \quad (7.71)$$

Suppose the oscillator fails when its stiffness drops to  $k_{m+1}$  during a reference time interval  $[0, \tau]$ , so that the probability of failure in  $\tau$  is

$$P_f(\tau) = P\left(\sum_{r=1}^m \Gamma_i \leq \tau\right) = P(T_m \leq \tau). \quad (7.72)$$

Under our assumptions, the random times  $\{\Gamma_r\}$  are independent random variables so that the characteristic function of the time to failure  $T_m$  is

$$\varphi_{T_m}(u) = E[e^{iuT_m}] = \prod_{r=1}^m E[e^{iu\Gamma_r}] = \prod_{r=1}^m \frac{i\mu_r(x_{\text{cr}})}{u + i\mu_r(x_{\text{cr}})} \quad (7.73)$$

since  $\Gamma_r$  is an exponential random variable with expectation  $1/\mu_r(x_{\text{cr}})$  ([43], Sect. 26). Since  $T_m$  is positive, its distribution can be calculated from ([44], Theorem 3.2.1)

$$\begin{aligned} F_{T_m}(t) = P(T_m \leq t) &= \lim_{\bar{u} \rightarrow \infty} \frac{1}{2\pi} \int_{-\bar{u}}^{\bar{u}} \frac{1 - e^{-iut}}{iu} \varphi_{T_m}(u) du \\ &= \lim_{\bar{u} \rightarrow \infty} \frac{1}{2\pi} \int_{-\bar{u}}^{\bar{u}} \frac{1 - e^{-iut}}{u} \prod_{r=1}^m \frac{\mu_r(x_{\text{cr}})}{u + i\mu_r(x_{\text{cr}})} du, \end{aligned} \quad (7.74)$$

or  $F_{T_m}(t) = \int_0^t f_{T_m}(s)ds$ , where

$$f_{T_m}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i u \tau} \varphi_{T_m}(u) du \quad (7.75)$$

is the density of  $T_m$ . The probability of failure in  $[0, \tau]$  is  $P_f(\tau) = F_{T_m}(\tau)$ .

The numerical results in the following figures are for  $M = 1$ ,  $C = \pi/10$ ,  $v_1 = \pi$ ,  $v_2 = 0.9\pi$ ,  $v_3 = 0.8\pi$ ,  $v_4 = 0.7\pi$ ,  $v_5 = 0.6\pi$ ,  $k_r = v_r^2$ , and  $\zeta_r = C/(2v_r)$ . The noise intensity and the critical threshold are  $g_0 = 1$  and  $x_{cr} = 1.5$ . The oscillator fails when its stiffness drops to  $k_5 = v_5^2 = (0.6\pi)^2$ . Other failure conditions can be considered. The average residence times in the first four damage states are  $E[\Gamma_1] = 4.3949$ ,  $E[\Gamma_2] = 3.6859$ ,  $E[\Gamma_3] = 3.2240$ , and  $E[\Gamma_4] = 2.9508$ . Figure 7.14 shows with solid and dotted lines the real and the imaginary parts of the characteristic functions of  $\Gamma_r$ ,  $r = 1, \dots, 4$ . The left and right panels in Fig. 7.15 show the density of failure time  $T_m$  and the failure probability  $P_f(\tau)$ .  $\diamond$

### 7.4.5 State Augmentation

The theory of stochastic differential equations can be applied to find properties of the solutions of both linear and nonlinear differential equations with random coefficients and input under some mild conditions, as has already been mentioned at the end of Sect. 7.4.1. For example, properties of  $X(t)$  in Example 7.26 can be obtained from those of the bivariate diffusion process with coordinates  $(X_1(t), X_2(t))$  defined by

$$\begin{cases} dX_1(t) = -X_2(t) X_1(t) dt + dB(t) \\ dX_2(t) = 0, \end{cases} \quad (7.76)$$

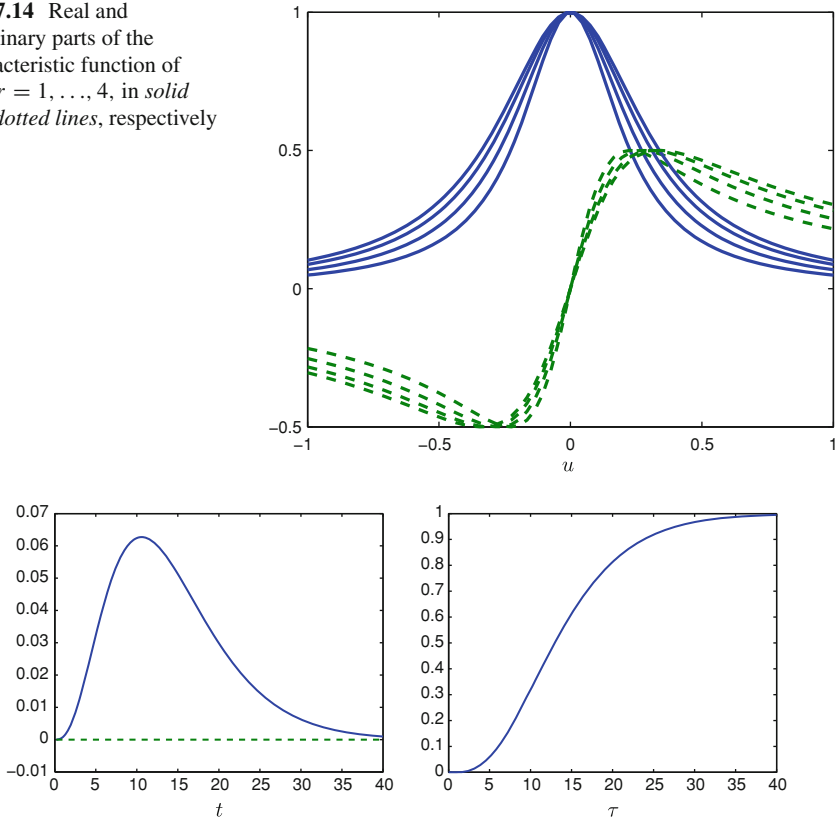
and initial conditions  $(X_1(0) \sim N(\mu_0, \gamma_0), X_2(0) = A)$ , where  $X_1(0)$ ,  $A$ , and  $B(t)$  are mutually independent. The drift and diffusion coefficients in (7.76) satisfy the uniform Lipschitz conditions in (5.29), but are not bounded so that Theorem 5.7 cannot be applied directly. Arguments as in Example 5.11 can be used to apply this theorem and conclude that (7.76) has a strong solution that is unique in the strong sense.

The assumptions that the driving noise is a Gaussian white noise and  $A$  is time invariant are not essential. For example, suppose  $A$  is a time-variant random coefficient and  $dB(t)$  is replaced by  $Y(t)dt$  defined by  $dA(t) = \alpha_1(A(t))dt + \alpha_2(A(t))dB_1(t)$  and  $dY(t) = \beta_1(Y(t))dt + \beta_2(Y(t))dB_2(t)$ , respectively, where  $B_1$  and  $B_2$  are mutually independent standard Brownian motions that are independent of  $B$ . The  $\mathbb{R}^3$ -valued stochastic process  $(X_1 = X, X_2 = A, X_3 = Y)$  defined by

$$\begin{cases} dX_1(t) = -X_2(t)X_1(t)dt + X_3(t)dt \\ dX_2(t) = \alpha_1(X_2(t))dt + \alpha_2(X_2(t))dB_1(t), \\ dX_3(t) = \beta_1(X_3(t))dt + \beta_2(X_3(t))dB_2(t) \end{cases} \quad (7.77)$$



**Fig. 7.14** Real and imaginary parts of the characteristic function of  $\Gamma_r$ ,  $r = 1, \dots, 4$ , in *solid* and *dotted lines*, respectively



**Fig. 7.15** Density of failure time  $T_m$  (left panel) and failure probability  $P_f(\tau)$  for  $x_{cr} = 1.5$  (right panel)

is a diffusion process. If the drift and the diffusion coefficients of (7.77) satisfy the conditions of Theorem 5.7 or 5.8, this equation has a unique solution.

While conceptually attractive, the state augmentation method has a limited use in applications since augmented equations have higher dimension than that of original equations and, usually, are difficult to solve. For example, the augmented equation for a linear system with random coefficients is a nonlinear stochastic differential equation, as illustrated by (7.76).

*Example 7.29* The moments  $\mu(p, q; t) = E[X_1(t)^p X_2(t)^q]$  of the bivariate diffusion process in (7.76) satisfy the ordinary differential equation

$$\dot{\mu}(p, q; t) = -p\mu(p, q + 1; t) + \frac{p(p-1)}{2}\mu(p-2, q; t)$$

with appropriate initial conditions. These equations cannot be solved exactly since they are not closed.

The density  $f(x_1, x_2; t)$  of  $(X_1(t), X_2(t))$  satisfies the Fokker–Planck equation

$$\frac{\partial f(x_1, x_2; t)}{\partial t} = \frac{\partial}{\partial x_1} \left[ x_1 x_2 f(x_1, x_2; t) + \frac{1}{2} \frac{\partial f(x_1, x_2; t)}{\partial x_1} \right]$$

so that, under the assumption  $A > 0$  a.s.,  $f(x_1, x_2; t)$  converges as  $t \rightarrow \infty$  to a time-invariant density  $f_s(x_1, x_2)$  defined by

$$\frac{\partial}{\partial x_1} \left[ x_1 x_2 f_s(x_1, x_2) + \frac{1}{2} \frac{\partial f_s(x_1, x_2)}{\partial x_1} \right] = 0.$$

The solutions of both the transient and stationary Fokker–Planck equations are difficult when dealing with realistic systems since the state vectors of these systems have usually large dimensions. For the simple case considered here, the stationary density of  $(X_1(t), X_2(t))$  has the expression

$$f_s(x_1, x_2) = \sqrt{x_2/\pi} \exp(-x_1^2 x_2) f_a(x_2)$$

where  $f_a$  denotes the density of  $A$ . Additional examples and considerations on the state augmentation method can be found elsewhere ([4], Sect. 9.2.4).  $\diamond$

*Proof* For large times, we have  $X(t) \mid A \sim N(0, 1/(2A))$ , so that its density is  $f_s(x_1 \mid x_2) = \sqrt{x_2/\pi} \exp(-x_1^2 x_2)$  and  $f_s(x_1, x_2) = f_s(x_1 \mid x_2) f_a(x_2)$ . Straight-forward calculations show that  $f_s(x_1, x_2)$  so defined satisfies the above stationary Fokker–Planck equation.  $\blacktriangle$

*Example 7.30* The displacement  $X(t)$  of a simple oscillator with unit mass, damping  $C > 0$  and stiffness  $K > 0$  subjected an harmonic action satisfies the differential equation  $\ddot{X}(t) + C\dot{X}(t) + KX(t) = \alpha \sin(\nu t)$ . It is assumed that  $C > 0$  and  $K > 0$  are independent random variables and  $\alpha, \nu > 0$  are some constants. The augmented  $\mathbb{R}^4$ -valued process  $Z(t)$  with coordinates  $(Z_1 = X, Z_2 = \dot{X}, Z_3 = C, Z_4 = K)$  is defined by the differential equation

$$dZ(t) = \begin{bmatrix} Z_2(t) \\ -Z_4(t)Z_1(t) - Z_3(t)Z_2(t) + \alpha \sin(\nu t) \\ 0 \\ 0 \end{bmatrix} dt = h(Z(t), t)$$

so that its density  $f(z; t)$  satisfies

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial z_1} (z_2 f) - \frac{\partial}{\partial z_2} [(-z_4 z_1 - z_3 z_2 + \alpha \sin(\nu t)) f],$$

known as the Liouville equation ([4], Sect. 9.2.5).  $\diamond$

*Proof* Let  $\varphi(u; t) = E[\exp(iu'Z(t))]$  denote the characteristic function of  $Z(t)$  so that  $f(z; t) = \int_{\mathbb{R}^4} \exp(-iu'z) \varphi(u; t) du / (2\pi)^4$ . The time derivative of  $\varphi(u; t)$  is

$$\frac{\partial \varphi}{\partial t} = i \sum_{k=1}^4 u_k E[\dot{Z}_k(t) e^{iu'Z(t)}] = i \sum_{k=1}^4 u_k E[h_k(Z(t), t) e^{iu'Z(t)}].$$

The Fourier transforms of the left and the right sides of this equation are  $\partial f / \partial t$  and  $-\sum_{k=1}^4 \int_{\mathbb{R}^4} [\partial(f h_k) / \partial z_k] \exp(iu'z) dz$ , respectively. Integration by parts of the latter term gives the right side of the Liouville equation.  $\blacktriangle$

### 7.4.6 Stochastic Reduced Order Models

Properties of stochastic reduced order models (SROMs) are discussed in Sect. A.3. These models are simple random elements, that is, elements that have a finite number  $m$  of samples that may not be equally likely. Optimization algorithms can be used to select the samples of SROMs and their probabilities. SROMs are used to approximate target random elements in the definition of stochastic differential equations and construct simple representations for the solutions of these equations.

*Example 7.31* Let  $X(t)$  be the solution of

$$\dot{X}(t) + AX(t) = Y(t), \quad t \in I = [0, \tau], \quad (7.78)$$

with  $A \sim U(a_1, a_2)$ ,  $0 < a_1 < a_2 < \infty$  and  $Y(t) = \cos(vt)$ . This equation is also solved by the stochastic Galerkin and stochastic collocation methods in Examples 7.33 and 7.35. Consider a SROM  $\tilde{A}$  for  $A$  with samples  $a^{(k)} = a_1 + (k-1)\Delta a$ ,  $k = 1, \dots, m$ ,  $m \geq 3$ , that are equally spaced at  $\Delta a = (a_2 - a_1)/(m-1)$ . The probabilities of these samples are selected to be  $p_1 = p_m = 1/(2(m-1))$  and  $p_k = 1/(m-1)$ ,  $k = 2, \dots, m-1$ . Note that  $\tilde{A}$  has been constructed by heuristic considerations, and may not be optimal.

The solution of  $\dot{X}(t) + AX(t) = \cos(vt)$ ,  $t \geq 0$ , with  $X(0) = 0$  and  $A$  set equal to  $a^{(k)}$ ,  $k = 1, \dots, m$ , is

$$X^{(k)}(t) = \frac{1}{(a^{(k)})^2 + v^2} \left( a^{(k)} \cos(vt) + v \sin(vt) - a^{(k)} \exp(-a^{(k)}t) \right). \quad (7.79)$$

The functions  $\{X^{(k)}(t)\}$  and their probabilities  $\{p_k\}$  of the samples of  $\tilde{A}$  define a SROM  $\tilde{X}(t)$  for  $X(t)$ , that is used to approximate properties of  $X(t)$ . For example, moments of order  $r$  of  $\tilde{X}(t)$  are given by

$$E[\tilde{X}(t)^r] = \sum_{k=1}^m p_k (X^{(k)}(t))^r, \quad t \geq 0. \quad (7.80)$$

The plots in the following two figures are for  $[a_1, a_2] = [1, 5]$  and  $v = 5$ . Figure 7.16 shows Monte Carlo estimates of  $E[X(t)^2]$  (left panel) and  $E[X(t)^4]$  (right panel) based on 1000 independent samples of  $X(t)$ . Approximations of these moments based

on a SROM of  $A$  with  $m = 5$  are indistinguishable from Monte Carlo estimates at the figure scale. The thin solid lines in Fig. 7.17 are Monte Carlo estimates of  $E[X(t)^4]$  based on sets of five independent samples of  $X(t)$ . The heavy solid line is the Monte Carlo estimate of  $E[X(t)^4]$  in Fig. 7.16 (right panel). In contrast to SROM-based estimates, Monte Carlo estimates of  $E[X(t)^4]$  based on  $m = 5$  samples are unstable, and can be inaccurate.  $\diamond$

*Example 7.32* Let  $X(t)$  be the solution of (7.78) with  $A = 1$  and  $Y(t) = \cos(\Theta t)$ ,  $\Theta \sim U(v_1, v_2)$ ,  $0 < v_1 < v_2 < \infty$ . This equation is also solved in Examples 7.34 and 7.36 by the stochastic Galerkin and stochastic collocation methods. Let  $\tilde{\Theta}$  with samples  $v^{(k)} = v_1 + (k - 1)\Delta v$ ,  $k = 1, \dots, m$ ,  $\Delta v = (v_2 - v_1)/(m - 1)$ ,  $m \geq 3$ , and probabilities  $p_1 = p_m = 1/(2(m - 1))$  and  $p_k = 1/(m - 1)$ ,  $k = 2, \dots, m - 1$ , be a SROM for  $\Theta$ . The solutions  $X^{(k)}(t)$  of  $\dot{X}(t) + X(t) = \cos(\Theta t)$ ,  $t \geq 0$ , for  $\Theta = v^{(k)}$  and  $X(0) = 0$  are

$$X^{(k)}(t) = \frac{1}{(1 + (v^{(k)})^2)} \left( \cos(v^{(k)}t) + v^{(k)} \sin(v^{(k)}t) - \exp(-t) \right), \quad (7.81)$$

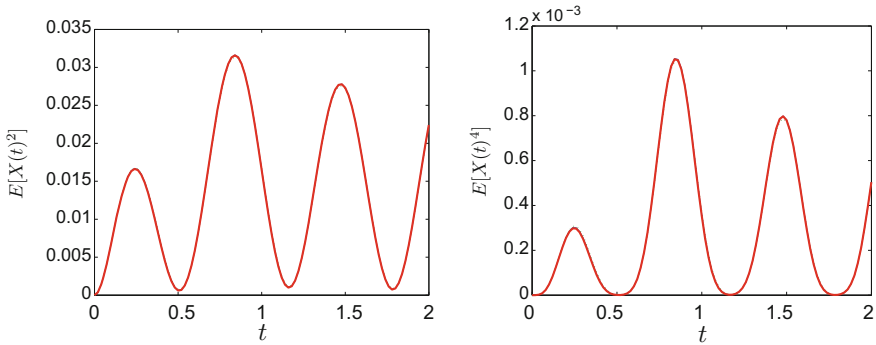
so that  $\tilde{X}(t)$  with samples  $\{X^{(k)}(t)\}$  and probabilities  $\{p_k\}$  is a SROM for  $X(t)$ . Properties of  $\tilde{X}(t)$  can be calculated simply, as illustrated in (7.80).

Numerical results in the following figures are for  $[v_1, v_2] = [1, 5]$  and several values of  $m$ . The solid lines in Fig. 7.18 are Monte Carlo estimates of  $E[X(t)^2]$  and  $E[X(t)^4]$  based on 1000 independent samples of  $X(t)$ . The dotted lines are approximations of these moments based on SROMs with  $m = 10$  (left panels) and  $m = 30$  (right panels). There is a notable improvement in the quality of the approximations  $E[\tilde{X}(t)^r]$  for  $E[X(t)^r]$  as the model size is increased from  $m = 10$  to  $m = 30$ . The approximations  $E[\tilde{X}(t)^r]$  corresponding to a SROM with  $m = 30$  trace the target moments over the entire time range. The heavy solid line in Fig. 7.19 is the Monte Carlo estimate in the right panels of Fig. 7.18. The thin solid lines in the figure are Monte Carlo estimates of  $E[X(t)^4]$  based on sets of 30 independent samples of  $X(t)$ . These estimates are unstable and have notable errors, in contrast to SROM-based approximations using  $m = 30$  samples.  $\diamond$

### 7.4.7 Stochastic Galerkin Method

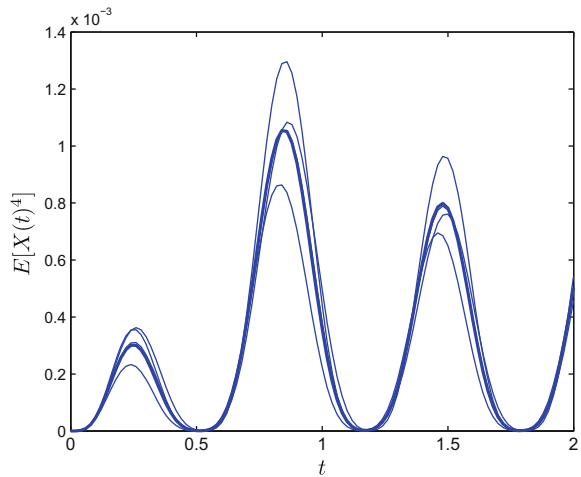
We consider stochastic differential equations of the type in (7.1) and (7.2). If  $X(t)$  is the solution of (7.1), the random functions in the definition of the drift and diffusion of this equation need to be approximated by parametric models, that is, deterministic functions of time depending on a finite number of random variables. This approximation, referred to as the discretization of the probability space, is examined to some extent in Sect. 9.4.6.1.

Galerkin solutions for stochastic differential equations with random coefficients are weak solutions for these equations corresponding to specified spaces of trial functions. We have seen in Examples 7.24 and 7.25 how the Babuška-Lax-Milgram



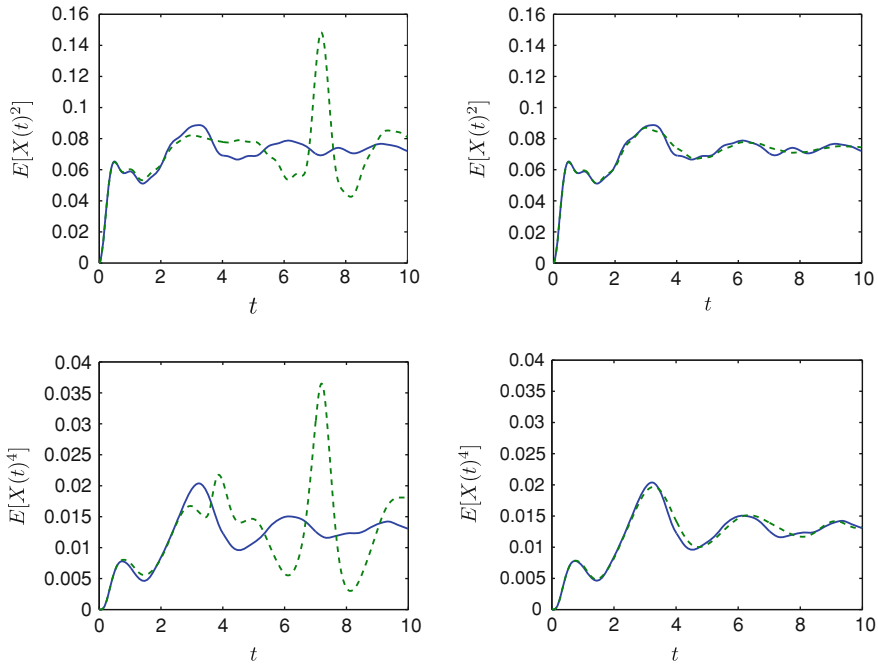
**Fig. 7.16** Monte Carlo estimates of  $E[X(t)^2]$  (left panel) and  $E[X(t)^4]$  (right panel) and SRM-based approximations for  $m=5$

**Fig. 7.17** Monte Carlo estimates of  $E[X(t)^4]$  based on 1000 independent samples of  $X(t)$  (heavy solid line) and on sets of five independent samples of  $X(t)$  (thin solid lines)



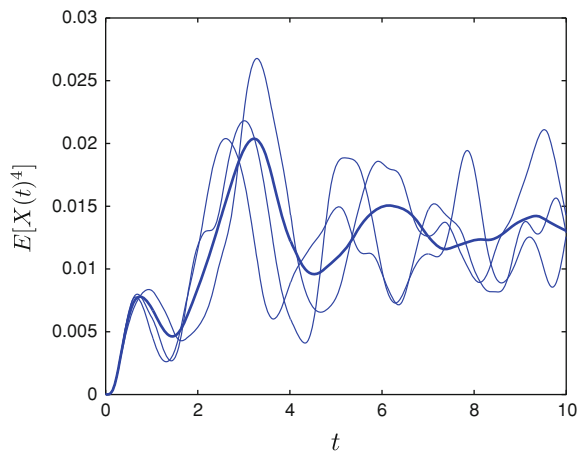
theorem can be applied to prove the existence and the uniqueness of weak solutions for stochastic differential equations with random coefficients. The following examples show how the stochastic Galerkin method can be implemented to calculate weak solutions for this class of stochastic equations.

**Example 7.33** Let solve the stochastic differential equation in Example 7.31 by the stochastic Galerkin method. We have seen in Example 7.24 that this equation admits unique weak solutions. Since  $A \sim U(a_1, a_2)$ ,  $0 < a_1 < a_2 < \infty$ , has bounded support, we expand both the random coefficients  $A$  and the solution  $X(t)$  in the subspace spanned by Legendre polynomials  $\{\phi_i(U), i = 0, 1, \dots, m\}$  up to a degree  $m \geq 1$ , that is,



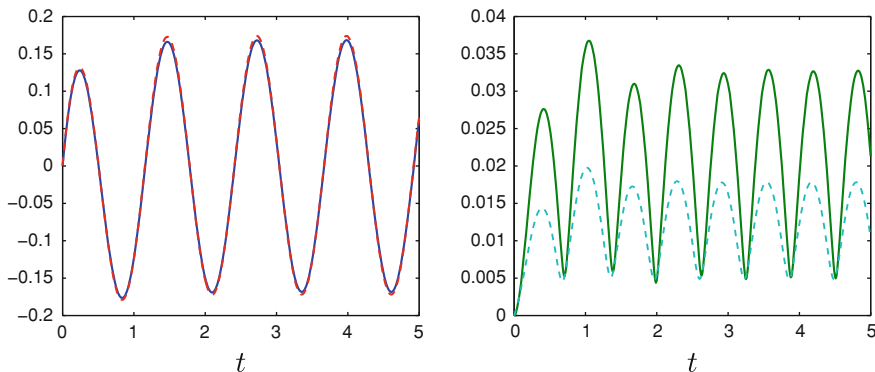
**Fig. 7.18** Monte Carlo estimates of  $E[X(t)^r]$  based on 1000 independent samples of  $X(t)$  (solid lines) and approximations of  $E[X(t)^r]$  (dash lines) based on SROMs with  $m = 10$  (left panels) and  $m = 30$  (right panels)

**Fig. 7.19** Monte Carlo estimates of  $E[X(t)^4]$  based on 1000 independent samples of  $X(t)$  (heavy solid line) and corresponding estimates based on sets of 30 independent samples of  $X(t)$  (thin solid lines)



$$A = \frac{a_1 + a_2}{2} + \frac{a_2 - a_1}{4} \phi_1(U)$$

$$X(t) \simeq \tilde{X}(t) = \sum_{i=0}^m \beta_i(t) \phi_i(U) \quad , \quad (7.82)$$



**Fig. 7.20** Means (*left panel*) and standard deviations (*right panel*) of  $X(t)$  and  $\tilde{X}(t)$  for  $m = 4$

where  $U \sim U(-1, 1)$ . With these representations, the defining equation for  $X(t)$  becomes

$$\sum_{i=0}^m \dot{\beta}_i(t) \phi_i(U) + \left( \frac{a_1 + a_2}{2} + \frac{a_2 - a_1}{4} \phi_1(U) \right) \sum_{i=0}^m \beta_i(t) \phi_i(U) = \cos(vt), \quad (7.83)$$

and has projections

$$\begin{aligned} & \sum_{i=0}^m \dot{\beta}_i(t) E[\phi_i(U) \phi_j(U)] + \frac{a_1 + a_2}{2} \sum_{i=0}^m \beta_i(t) E[\phi_i(U) \phi_j(U)] \\ & + \frac{a_2 - a_1}{4} E[\phi_1(U) \phi_i(U) \phi_j(U)] = \cos(vt) E[\phi_j(U)], \quad j = 0, 1, \dots, m, \end{aligned} \quad (7.84)$$

on  $\phi_j(U)$ . The coefficients  $\{\beta_i(t)\}$  in the representation of  $\tilde{X}(t)$  satisfy a system of coupled linear differential equations. The solution of this system of equations and the expression of  $\tilde{X}(t)$  can be used to find properties of  $X(t)$  approximately.

Figure 7.20 shows estimates of the means (*left panel*) and standard deviations (*right panel*) of  $X(t)$  and  $\tilde{X}(t)$  with solid and dotted lines, respectively. The estimates are based on 1000 independent samples of these processes, where  $\tilde{X}(t)$  is given by (7.82) with  $m = 4$ . The first moment of the approximate solution  $\tilde{X}(t)$  traces closely the corresponding moment of  $X(t)$ .  $\diamond$

*Proof* It is argued correctly that Legendre rather than Hermite polynomials should be used to represent random coefficients with bounded support, such as  $A$  in this illustration. However, Legendre polynomials may not be adequate for solutions of stochastic equations whose probability laws may differ significantly from those of their coefficients.

We consider trial functions  $\phi(t; U) = \sum_{i=0}^m \beta_i(t) \phi_i(U)$ , where  $\phi(0; U) = 0$  and  $\beta_i(t)$  are real-valued functions of time. The corresponding members of  $\mathcal{V}$  and

$\mathcal{W}$  in (7.46) are such that  $E[\int_0^\tau \phi(t; U)^2 dt] < \infty$  and  $E[\int_0^\tau \dot{\phi}(t; U)^2 dt] < \infty$ , respectively. The representation in (7.82) is consistent with the assumption that  $X(t)$  is a member of  $\mathcal{W}$ .

The Legendre polynomials are defined by  $\phi_i(x) = d^i((x^2 - 1)^i)/dx^i$ ,  $i = 0, 1, \dots$ , and can be obtained from the recurrence formula

$$\phi_{i+1}(x) = 2(2i + 1)x\phi_i(x) - 4i^2\phi_{i-1}(x), \quad x \in [-1, 1], \quad (7.85)$$

for  $i \geq 2$  with  $\phi_0(x) = 1$  and  $\phi_1(x) = 2x$  (Sect. B.6). These polynomials have norms  $\|\phi_i\|^2 = E[\phi_i(U)^2] = (i!)^2 2^{2i}/(2i + 1)$ . ▲

*Example 7.34* Let solve the stochastic differential equation in Example 7.32 by the stochastic Galerkin method. The construction of the Galerkin solution is simple in this case since the uncertainty is in the input and the differential equation  $\dot{X}(t) + AX(t) = Y(t)$ ,  $Y(t) = \cos(\Theta t)$ , is linear. Since  $\Theta$  takes values in a bounded interval, we view the input  $Y(t)$  as a member of the space spanned by Legendre polynomials  $\{\phi_i(U), i = 0, 1, \dots, m\}$ ,  $U \sim U(-1, 1)$ , up to a degree  $m \geq 1$ , that is,

$$Y(t) \simeq \tilde{Y}(t) = \sum_{i=0}^m \alpha_i(t) \phi_i(U), \quad (7.86)$$

where  $\alpha_i(t) = E[\cos(\Theta t)\phi_i(U)]/E[\phi_i(U)^2]$ . The solution  $X(t)$  for  $Y(t)$  in (7.86) gives  $X(t) \simeq \tilde{X}(t) = \sum_{i=0}^m \beta_i(t) \phi_i(U)$ , where  $\beta_i(t) = \int_0^t \exp(-(t-s))\alpha_i(s)ds$ . Properties of  $X(t)$  can be approximated by those of  $\tilde{X}(t)$ , for example,  $E[X(t)^r] \simeq E[\tilde{X}(t)^r] = E[(\sum_{i=0}^m \beta_i(t)\phi_i(U))^r]$  and can be estimated simply and efficiently from samples of  $U$ . Numerical results are for  $A = 1$ .

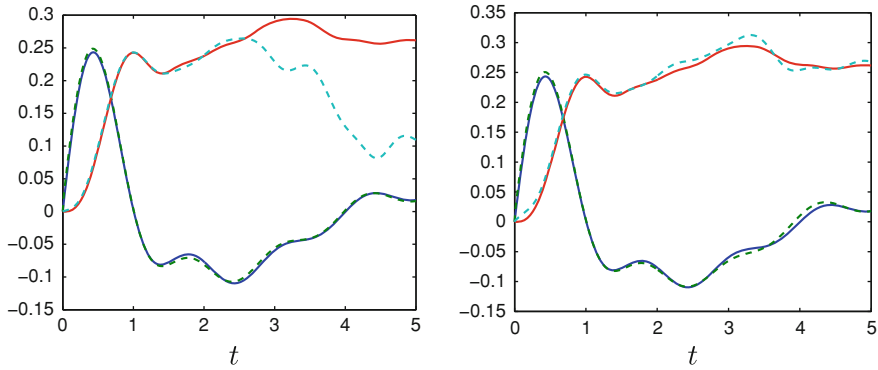
The solid and dotted lines in Fig. 7.21 are estimates of the first two moments of  $X(t)$  and  $\tilde{X}(t)$  for  $m = 4$  (left panel) and  $m = 9$  (right panel) obtained from 1000 independent samples of these processes. The plots that take only positive values and positive/negative values are standard deviations and expectations, respectively. There is a notable improvement in the quality of the approximate moments of  $\tilde{X}(t)$  as  $m$  is increased. This is consistent with the fact that  $\tilde{Y}(t)$  converges in m.s. to  $Y(t)$  as  $m \rightarrow \infty$ . Yet, the accuracy of higher order moments of  $\tilde{X}(t)$  improves much slower with  $m$ . For example, the maxima of  $(E[X(t)^4] - E[\tilde{X}(t)^4])/E[X(t)^4]$  in percentages are 90.81%, 25%, and 22% for  $m = 4, 9$ , and 19, respectively. ◇

*Proof* The coefficients  $\alpha_i(t)$  in the representation of  $\tilde{Y}(t)$  are obtained by projecting this process on the coordinates of the selected basis, that is, by imposing the condition  $E[\cos(\Theta t)\phi_j(U)] = E[\sum_{i=0}^m \alpha_i(t)\phi_i(U)\phi_j(U)]$ . The representation of  $\tilde{X}(t)$  and the defining equation for  $X(t)$  give

$$\sum_{i=0}^m \dot{\beta}_i(t) \phi_i(U) + \sum_{i=0}^m \beta_i(t) \phi_i(U) = \sum_{i=0}^m \alpha_i(t) \phi_i(U).$$

The projection of this equation on  $\phi_j(U)$ ,  $j = 0, 1, \dots, m$ , gives the differential equation  $\dot{\beta}_j(t) + \beta_j(t) = \alpha_j(t)$ ,  $j = 0, 1, \dots, m$ , with initial condition  $\beta_j(0) = 0$





**Fig. 7.21** Means and standard deviations of  $X(t)$  and  $\tilde{X}(t)$  for  $m = 4$  (left panel) and  $m = 9$  (right panel)

following from  $X(0) = 0$ . Note that, in contrast to the problem in Example 7.33, the equations defining the coefficients  $\{\beta_i(t)\}$  of the approximate solutions are decoupled. Moreover, the equations for  $\{\beta_i(t)\}$  delivered by the weak formulation coincide with those obtained by direct calculations using the linearity of the defining equation for  $X(t)$  and the representation of  $Y(t)$ . ▲

### 7.4.8 Stochastic Collocation Method

Let  $X(t)$ ,  $t \in [0, \tau]$ , be an  $\mathbb{R}^d$ -valued stochastic process satisfying a stochastic differential equation of the type in (7.2) with coefficients and/or input depending on a random vector  $\Theta$ . If the coefficients and/or input are random functions, they need to be approximated by parametric models, that is, finite sums of specified deterministic functions of time with random coefficients. We have seen that similar representations are needed to implement the stochastic Galerkin method.

For simplicity,  $X(t)$ ,  $t \in [0, \tau]$ , is assumed to be a real-valued stochastic process defined by a stochastic differential equation with random coefficients and input. Suppose the random functions in the definition of this equation are parametric and depend on a random vector  $\Theta$  defined on a probability space  $(\Omega, \mathcal{F}, P)$ . The collocation solutions for this equations have the form

$$\tilde{X}(t, \sigma) = \sum_{i=1}^m X^{(i)}(t) \ell_i(\sigma), \quad t \in [0, \tau], \quad \sigma \in \Gamma, \quad (7.87)$$

where  $\Gamma = \Theta(\Omega) \subset \mathbb{R}^d$  is the range of  $\Theta$ ,  $\{\sigma^{(i)} \in \Gamma, i = 1, \dots, m\}$  denote collocation points, and  $X^{(i)}(t) = X(t, \sigma^{(i)})$  are solutions of the deterministic equations obtained from the stochastic differential equation with  $\Theta$  set equal to  $\sigma^{(i)} \in \Gamma$ ,  $i = 1, \dots, m$ . As for the stochastic Galerkin method, it is assumed that  $\Gamma$

is a bounded set in  $\mathbb{R}^d$  and that the equation defining  $X(t)$  has solutions for almost all  $(t, y) \in [0, \tau] \times \Gamma$ . A broad range of theoretical considerations on the stochastic collocation method and practical aspects related to the implementation of this method can be found, for example, in [45–47].

**Theorem 7.18** *If  $X(t, \sigma)$  is a continuous function of  $\sigma \in \Gamma$  for a fixed  $t \in [0, \tau]$ , there exists an algebraic polynomial  $p(t, \cdot) : \Gamma \rightarrow \mathbb{R}$  such that  $\|X(t, \cdot) - p(t, \cdot)\|_\infty \leq \varepsilon$  for arbitrary  $\varepsilon > 0$ .*

*Proof* The polynomial  $p$  in the theorem is the collocation approximation  $\tilde{X}(t, \sigma)$  at time  $t$ . The proof follows from properties of polynomial approximations (Theorems 8.10 and 8.15).

Note that the random variables  $X(t)$  are bounded with probability 1 for each  $t$  since  $X(t, \cdot) : \Gamma \rightarrow \mathbb{R}$  is assumed to be continuous and  $\Gamma$  is bounded. The degree of the polynomial  $p$  satisfying  $\|X(t, \cdot) - p(t, \cdot)\|_\infty \leq \varepsilon$  depends on  $t$  since the properties of  $X(t, \cdot)$  change in time. If  $X(t, \sigma)$  is approximated by polynomials of  $\sigma$  with the same functional forms and time dependent coefficients, the accuracy of the resulting collocation solutions will not be uniform in  $t$ . ▲

**Theorem 7.19** *Under the assumptions in Theorem 7.18, the discrepancy between the correlations  $E[X(s)X(t)]$  and  $E[\tilde{X}(s)\tilde{X}(t)]$  can be made as small as desired for arbitrary but fixed times  $s$  and  $t$ .*

*Proof* For  $s, t \in [0, \tau]$  fixed we have

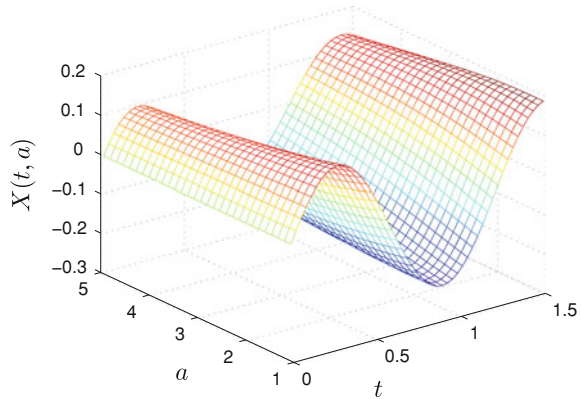
$$\begin{aligned} |E[X(s)X(t)] - E[\tilde{X}(s)\tilde{X}(t)]| &= |E[X(s)X(t) - \tilde{X}(s)\tilde{X}(t)]| \\ &\leq E[|X(s)X(t) - \tilde{X}(s)\tilde{X}(t)|] \leq E[|X(s)||X(t) - \tilde{X}(t)| + |X(s) - \tilde{X}(s)||\tilde{X}(t)|] \\ &\leq (E[X(s)^2]E[(X(t) - \tilde{X}(t))^2])^{1/2} + (E[(X(s) - \tilde{X}(s))^2]E[\tilde{X}(t)^2])^{1/2} \end{aligned}$$

so that  $|E[X(s)X(t)] - E[\tilde{X}(s)\tilde{X}(t)]|$  can be made smaller than  $\varepsilon > 0$  since  $E[X(s)^2]$  and  $E[\tilde{X}(t)^2]$  are bounded and it is possible to construct a polynomial approximation  $\tilde{X}$  whose error at  $s$  and  $t$  be smaller than  $\varepsilon/(2c)$ , where  $c > 0$  is a finite constant such that  $E[X(s)^2] \vee E[\tilde{X}(t)^2] \leq c$ .

Similar arguments can be used to show that  $|E[X(s)X(t)] - E[\tilde{X}(s)\tilde{X}(t)]| \leq \varepsilon$  holds at a finite number of times  $s$  and  $t$ . However, additional conditions are needed for  $E[\tilde{X}(s)\tilde{X}(t)]$  to approximate  $E[X(s)X(t)]$  within  $\varepsilon$  at all times  $s, t \in [0, \tau]$ . ▲

**Example 7.35** Let  $X(t)$  be the solution of (7.78), that is, this process satisfies the equation  $\dot{X}(t) + AX(t) = Y(t)$ , with  $A \sim U(a_1, a_2)$ ,  $0 < a_1 < a_2 < \infty$  and  $Y(t) = \cos(vt)$ ,  $v > 0$ . We have solved this problem by the SROM and stochastic Galerkin methods in Examples 7.31 and 7.33. The collocation solution in (7.87) becomes

$$\tilde{X}(t, a) = \sum_{i=1}^m X^{(i)}(t) \ell_i(a) = \sum_{i=1}^m X^{(i)}(t) \prod_{j=1, j \neq i}^m \frac{a - a^{(j)}}{a^{(j)} - a^{(i)}}, \quad (7.88)$$

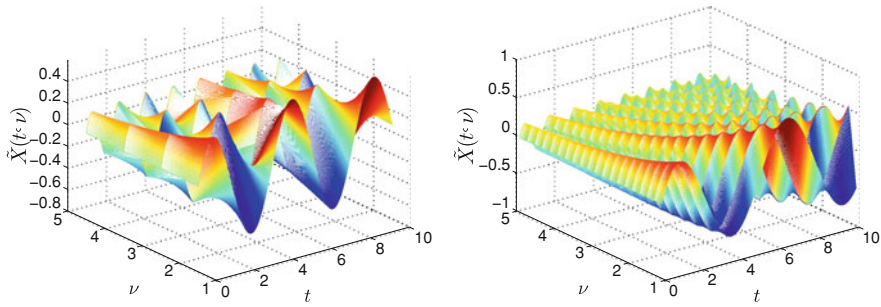
**Fig. 7.22** Mapping $(t, a) \mapsto X(t, a)$  for $\tau = 1, a_1 = 0.1, a_2 = 5,$   
and  $v = 5$ 

where  $a^{(i)} \in A(\Omega) = [a_1, a_2]$ ,  $i = 1, \dots, m$ , denote collocation points and  $X^{(i)}(t)$  is given by (7.79). Figure 7.22 shows the variation of the solution  $X(t, a)$  in  $[0, \tau] \times [a_1, a_2]$  for  $\tau = 1$ ,  $a_1 = 0.1$ ,  $a_2 = 5$ , and  $v = 5$ . Collocation solutions using  $m \geq 4$  equally spaced points in  $[a_1, a_2]$  are very accurate, in agreement with Theorem 7.18 since  $X$  is continuous in  $a \in [a_1, a_2]$  and varies slowly with this argument at all times.  $\diamond$

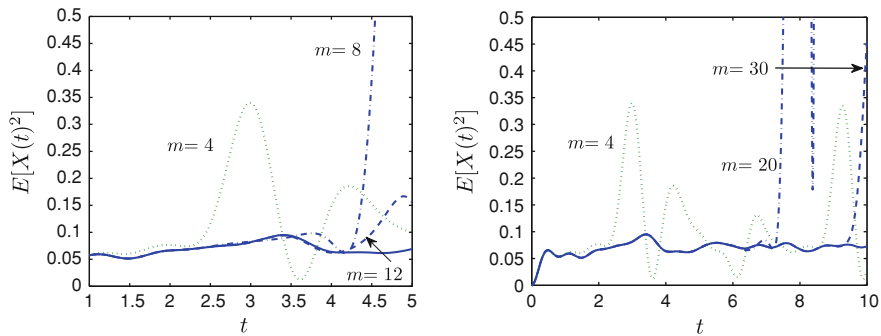
*Example 7.36* Let  $X(t)$  be the solution of (7.78), that is, the equation  $\dot{X}(t) + AX(t) = Y(t)$ , with  $A = 1$ ,  $Y(t) = \cos(\Theta t)$ ,  $\Theta \sim U(v_1, v_2)$ ,  $0 < v_1 < v_2 < \infty$ . We have solved this problem by the SROM and stochastic Galerkin methods in Examples 7.32 and 7.34. The collocation solution  $\tilde{X}(t, v)$  is given by (7.88) with  $(v, v^{(i)})$ ,  $v^{(i)} \in [v_1, v_2]$ , in place of  $(a, a^{(i)})$  and  $X^{(i)}(t)$  in (7.81).

The following numerical results are for  $[v_1, v_2] = [1, 5]$  and equally spaced collocation points in this interval. Figure 7.23 shows the mapping  $(t, v) \mapsto \tilde{X}(t, v)$  for  $m = 4$  (left panel) and  $m = 20$  (right panel). The plots suggest that a relatively large number of collocation points is needed to capture the dependence of  $\tilde{X}(t, v)$  on parameter  $v$ . The solid lines in Fig. 7.24 are estimates of  $E[X(t)^2]$  obtained from 1000 independent samples of  $X(t)$ . The other lines are moments  $E[\tilde{X}(t)^2]$  obtained from a collocation solution with several values on  $m$  for  $\tau = 5$  (left panel) and  $\tau = 10$  (right panel). The accuracy of  $E[\tilde{X}(t)^2]$  relative to Monte Carlo estimates of  $E[X(t)^2]$  depends on the number  $m$  of collocation points but does not necessarily improve with  $m$ , even for sets of refining collocation points. A sequence of collocation points is refining if the points for a collocation solution with  $m = m_1$  are included in those for a collocation solution with  $m = m_2 > m_1$ .  $\diamond$

The results in Fig. 7.24 are consistent with Theorem 7.18 stating that the accuracy of collocation solutions is not uniform in time. For example, the approximate moment  $E[\tilde{X}(t)^2]$  for  $m = 8$  is superior to that for  $m = 4$  in the time interval  $[0, 4]$  but not in  $[0, 5]$ . That collocation solutions may not improve with  $m$  poses notable difficulties in applications since  $m$  determines the computation effort.



**Fig. 7.23** Mapping  $(t, \nu) \mapsto \tilde{X}(t, \nu)$  for  $\tau = 10$ ,  $[v_1, v_2] = [1, 5]$ ,  $m = 4$  (left panel) and  $m = 20$  (right panel)



**Fig. 7.24** Monte Carlo estimates of  $E[X(t)^2]$  and approximations  $E[\tilde{X}(t)^2]$  corresponding to various values of  $m$  for  $\tau = 5$  (left panel) and  $\tau = 10$  (right panel)

**Theorem 7.20** *Under the conditions of Theorem 7.18, it is possible to construct a collocation solutions  $\tilde{X}(t)$  whose marginal and finite dimensional distributions are as closed as desired to the corresponding distributions of  $X(t)$ .*

*Proof* The discrepancy between the characteristic functions of  $X(t)$  and  $\tilde{X}(t)$  is

$$\begin{aligned}
 |E[e^{iuX(t)}] - E[e^{iu\tilde{X}(t)}]| &= |E[e^{iuX(t)}(1 - e^{iu(\tilde{X}(t) - X(t))})]| \\
 &\leq \left( E[|1 - e^{iu(\tilde{X}(t) - X(t))}|^2] \right)^{1/2} \\
 &= \left( E\left[ (1 - \cos(u(\tilde{X}(t) - X(t))))^2 + \sin^2(u(\tilde{X}(t) - X(t))) \right] \right)^{1/2} \\
 &\leq \sqrt{2}|u|(E[(\tilde{X}(t) - X(t))^2])^{1/2}
 \end{aligned}$$

by the Cauchy–Schwarz inequality, properties of the characteristic function, and the inequalities  $\sin^2(x) \leq x^2$  and  $(1 - \cos(x))^2 \leq x^2$ . For an arbitrary  $\varepsilon' > 0$  and  $0 \leq \bar{u} < \infty$ , there is a collocation approximation  $\tilde{X}$  such that  $\|X(t, \cdot) - \tilde{X}(t, \cdot)\|_\infty \leq$

$\varepsilon = \varepsilon' / (\sqrt{2}\bar{u})$  so that  $|E[e^{iuX(t)}] - E[e^{iu\tilde{X}(t)}]| \leq \varepsilon'$ . Since  $\bar{u}$  is arbitrary, we conclude that it is possible to construct a collocation solution  $\tilde{X}(t)$  such that the marginal characteristic functions of  $\tilde{X}(t)$  and  $X(t)$  are as close as desired. Similar arguments hold for the characteristic functions of the vectors  $(X(t_1), \dots, X(t_n))$  and  $(\tilde{X}(t_1), \dots, \tilde{X}(t_n))$ , where  $n \geq 1$  is an integer and  $(t_1, \dots, t_n)$  are points in  $[0, \tau]$ .  $\blacktriangle$

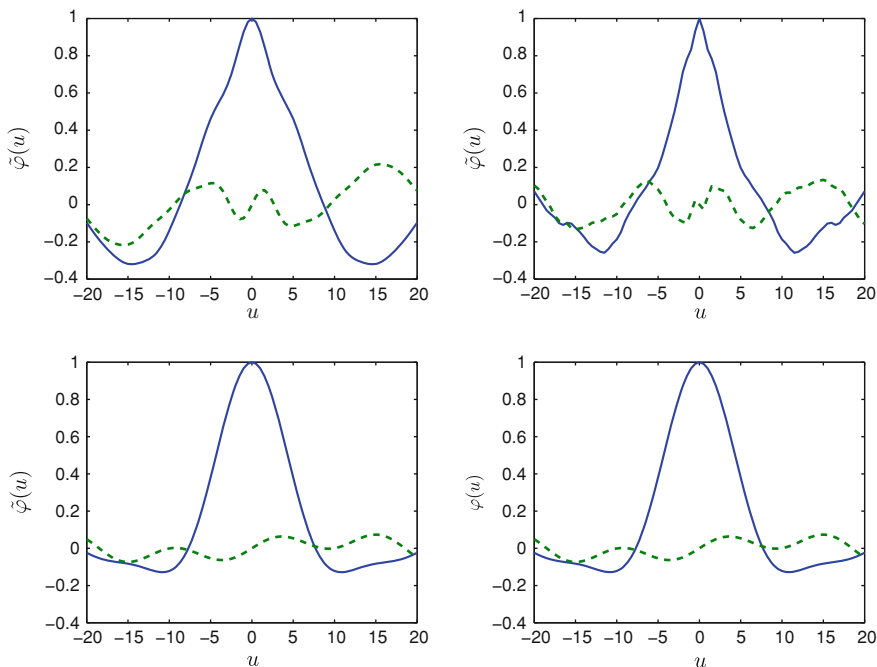
*Example 7.37* Let  $\varphi(u; t) = E[\exp(iuX(t))]$  and  $\tilde{\varphi}(u; t) = E[\exp(iu\tilde{X}(t))]$  denote the characteristic functions of  $X(t)$  and  $\tilde{X}(t)$  in Example 7.36. The top left, top right, and bottom left panels in Fig. 7.25 show characteristic functions  $\tilde{\varphi}(u; t)$  for  $m = 4$ ,  $m = 8$ , and  $m = 20$  at  $t = 4.5$ . The bottom right panel shows the characteristic function  $\varphi(u; t)$  at  $t = 4.5$ . The solid and dotted lines are the real and imaginary parts of the approximate and exact characteristic functions. Numerical values are for  $\Theta$  uniformly distributed in  $[v_1, v_2] = [1, 5]$ , as in Example 7.36. Monte Carlo estimates are based on 1000 independent samples.

The plots are consistent with the statement of Theorem 7.20 in the sense that it is possible to construct a collocation solution such that  $\tilde{\varphi}(u; t)$  be as close as desired to  $\varphi(u; t)$  at an arbitrary but fixed  $t$ , for example,  $\tilde{\varphi}(u; t)$  for  $m = 20$  and  $t = 4.5$ . The plots are also consistent with a comment in Example 7.36 that the accuracy of collocation solutions may not increase monotonically with  $m$ . For example, the discrepancy  $\max_{-5 \leq u \leq 5} |\varphi(u; t) - \tilde{\varphi}(u; t)|$  between  $\varphi(u; t)$  and  $\tilde{\varphi}(u; t)$  increases from 0.1886 for  $m = 4$  to 0.2350 for  $m = 8$ .  $\diamond$

We conclude this section with comments on similarities and differences between solutions of stochastic differential equations with random coefficients by the SROM, stochastic Galerkin, and stochastic collocation methods.

Solutions of stochastic differential equations by SROMs and stochastic collocation are constructed from deterministic solutions of these equations corresponding to samples of their random coefficients and driving noise processes. The stochastic collocation method requires that the random elements in the definition of stochastic equations be described by parametric models depending on a random vector  $\Theta$  such that  $\Theta(\Omega)$  is a bounded set. The selection of the number and the location of collocation points in  $\Theta(\Omega)$  relates to the computation effort and the accuracy of the polynomial representations in the probability space. The probability law of  $\Theta$  is not used to select collocation points. It is common to assume that  $\Theta$  has independent coordinates. SROM-based solutions do not require to approximate the random entries of stochastic equations by parametric models. However, the current version of ESROMs assumes that all random entries can be modeled by parametric models. The samples of the reduced order models used to approximate random coefficients and inputs as well as their probabilities are selected via optimization algorithms with objective functions quantifying discrepancies between features of the probability laws of target random elements and their SROMs.

There are no similarities between SROM- and Galerkin-based solutions. The stochastic Galerkin and collocation methods are similar in the sense that both methods use parametric models to describe approximately the random functions in the definition of stochastic equations. In contrast to the collocation method that construct the



**Fig. 7.25** Characteristic functions  $\tilde{\varphi}(u; t)$  for  $m = 4$  (top left panel),  $m = 8$  (top right panel), and  $m = 20$  (bottom left panel) and characteristic function  $\varphi(u; t)$  (bottom right panel) at  $t = 4.5$

solution of a stochastic equation by interpolating between solutions of this equation corresponding to collocation points, the stochastic Galerkin method selects a finite stochastic basis and assumes that the solution belongs to the subspace spanned by this basis. The coefficients of the solution are calculated from a system of equations obtained by projecting the proposed form of the solution on the members of the selected basis.

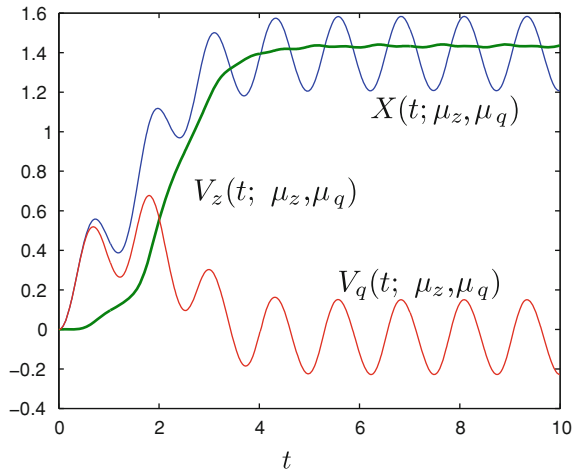
#### 7.4.9 Taylor, Perturbation, and Neumann Series

We consider stochastic differential equations with random coefficients and input that have small uncertainty. Examples illustrate the solution of this class of equations by the Taylor, perturbation, and Neumann series methods.

**Example 7.38** Let  $X(t)$  be the solution of  $(d/dt - \alpha + ZX(t)^2)X(t) = Q \sin(vt)$ ,  $t \geq 0$ , where  $(Z, Q)$  are uncorrelated random variables with finite variance, and  $\alpha, v > 0$  are constants. Let  $X(t; \mu_z, \mu_q)$ ,  $V_z(t; \mu_z, \mu_q)$ , and  $V_q(t; \mu_z, \mu_q)$  denote the solution  $X(t; Z, Q)$  of this equation and its partial derivatives with respect to  $Z$  and  $Q$

**Fig. 7.26** Functions

$X(t; \mu_z, \mu_q)$ ,  $V_z(t; \mu_z, \mu_q)$ ,  
and  $V_q(t; \mu_z, \mu_q)$  for  $\alpha =$   
1,  $\mu_z = -0.5$ ,  $\mu_q = 1$ , and  
 $v = 5$



evaluated at the mean  $(\mu_z, \mu_q)$  of  $(Z, Q)$ . The first order Taylor approximation of the solution  $X(t; Z, Q)$  viewed as a function of  $(Z, Q)$  is

$$X(t; Z, Q) \simeq X(t; \mu_z, \mu_q) + V_z(t; \mu_z, \mu_q)(Z - \mu_z) + V_q(t; \mu_z, \mu_q)(Q - \mu_q),$$

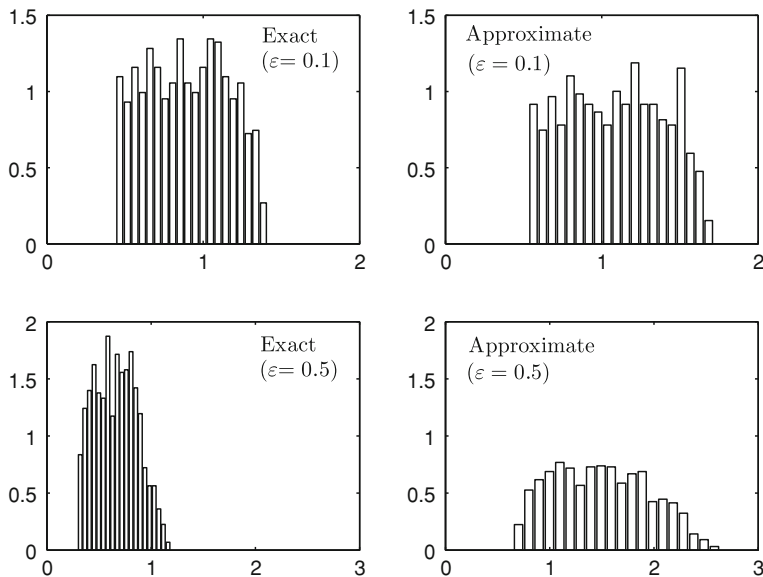
so that the approximate mean and variance functions are  $E[X(t)] \simeq X(t; \mu_z, \mu_q)$  and  $\text{Var}[X(t)] \simeq V_z(t; \mu_z, \mu_q)^2 \sigma_z^2 + V_q(t; \mu_z, \mu_q)^2 \sigma_q^2$ , where  $\sigma_z^2$  and  $\sigma_q^2$  denote the variances of  $Z$  and  $Q$ . Note that these moments depend only on the first two moments of  $Z$  and  $Q$ .

Figure 7.26 shows the deterministic functions  $X(t; \mu_z, \mu_q)$ ,  $V_z(t; \mu_z, \mu_q)$ , and  $V_q(t; \mu_z, \mu_q)$  for  $\alpha = 1$ ,  $\mu_z = -0.5$ ,  $\mu_q = 1$ ,  $v = 5$ , and  $X(0) = 0$ . We use these functions to calculate the mean and variance of  $X(t)$  approximately.  $\diamond$

*Proof* The approximate representation for  $X(t; Z, Q)$  consists of the first terms of the Taylor series expansion of  $X(t; Z, Q)$  view as a function of  $(Z, Q)$ . The deterministic functions  $X(t; \mu_z, \mu_q)$ ,  $V_z(t; \mu_z, \mu_q)$ , and  $V_q(t; \mu_z, \mu_q)$  are the solutions of the differential equation for  $X$  and of the derivatives of this equations with respect to  $Z$  and  $Q$  for  $(Z, Q)$  set equal to  $(\mu_z, \mu_q)$ . For example,  $V_q(t; \mu_z, \mu_q)$  is the solution of  $(d/dt - \alpha)V_q(t; \mu_z, \mu_q) - 3\mu_z X(t; \mu_z, \mu_q)^2 V_q(t; \mu_z, \mu_q) = \sin(vt)$ .  $\blacktriangle$

**Example 7.39** Let  $X$  be the solution of  $(d/dt + \beta + \varepsilon Y(t))X(t) = Q$ ,  $t \geq 0$ , where  $Y(t)$  is a stochastic process with mean zero,  $Q$  denotes a random variable, and  $\varepsilon$  is a small parameter. The first order perturbation solution is  $X(t) \simeq X^{(1)}(t) = X_0(t) + \varepsilon X_1(t)$ , where  $X_0(t) = (Q/\beta)(1 - e^{-\beta t})$ ,  $X_1(t) = -(Q/\beta)Z(t)e^{-\beta t}$ , and  $Z(t) = \int_0^t (e^{\beta s} - 1)Y(s)ds$ . Moments and other probabilistic properties of  $X(t)$  can be calculated approximately from this or higher order perturbation solutions.

Figure 7.27 shows histograms of  $X(t)$  and  $X^{(1)}(t)$  for  $X(0) = 0$ ,  $\beta = 1$ ,  $t = 5$ ,  $Q \sim U(0.5, 1.5)$ , and  $Y(t) = Y \sim U(0.5, 1.5)$  obtained from 1000 independent of these solutions under the assumption that  $Q$  and  $Y$  are independent random variables.



**Fig. 7.27** Histograms of the exact and first order perturbation solutions for  $X(0) = 0$ ,  $\beta = 1$ ,  $t = 5$ ,  $Q \sim U(0.5, 1.5)$ , and  $Y(t) = Y \sim U(0.5, 1.5)$

The histograms of  $X(t)$  and  $X^{(1)}(t)$  are similar for  $\varepsilon = 0.1$  but differ significantly for  $\varepsilon = 0.5$ .  $\diamond$

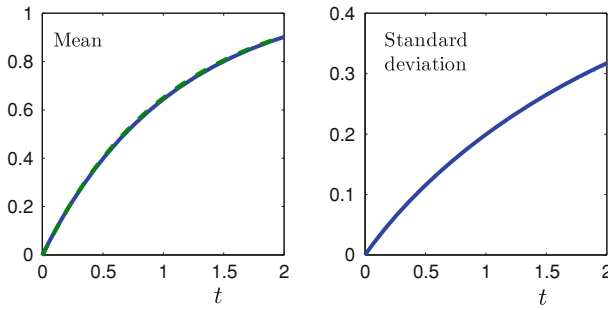
*Proof* The random functions  $X_0$  and  $X_1$  satisfy the differential equations  $\dot{X}_0(t) + \beta X_0(t) = Q$  with  $X_0(0) = 0$  and  $\dot{X}_1(t) + \beta X_1(t) = -Y(t)X_0(t)$  with  $X_1(0) = 0$ , so that they have the stated solutions. We have  $E[X(t)] = (E[Q]/\beta)(1 - e^{-\beta t})$  and  $E[X(s)X(t)] = E[(X_0(s) + \varepsilon X_1(s))(X_0(t) + \varepsilon X_1(t))] = E[X_0(s)X_0(t)] + \varepsilon E[X_0(s)X_1(t) + X_1(s)X_0(t)] + \varepsilon^2 E[X_1(s)X_1(t)]$ . Note that  $E[X(s)X(t)]$  would include two additional terms of order  $\varepsilon^2$  if based on higher order perturbations, the expectations  $E[X_0(s)X_2(t)]$  and  $E[X_2(s)X_0(t)]$ .  $\blacktriangle$

*Example 7.40* Let  $X(t)$  be the solution of  $(d/dt + \beta + Y)X(t) = Q$ ,  $t \in [0, \tau]$  where  $Y$  and  $Q$  are independent random variables and  $|Y|/\tau < 1$  a.s. The Neumann series for  $X(t)$  is absolutely and uniformly convergent a.s. in  $[0, \tau]$ , and

$$X(t) = f(t) - \lambda \int_0^\tau H(t, s; \lambda) f(s) ds, \quad (7.89)$$

where  $\lambda = -1$ ,  $f(t) = (Q/\beta)(1 - e^{-\beta t})$ ,  $H(t, s; \lambda) = -\sum_{r=1}^\infty \lambda^{r-1} K_r(t, s)$ , and the kernels  $K_r(t, s)$  can be obtained from  $K_r(t, s) = \int_0^\tau K(t, \sigma) K_{r-1}(\sigma, s) d\sigma$  for  $r \geq 2$  with  $K_1(t, s) = K(t, s) = 1(0 \leq s \leq t) Y \exp(-\beta(t-s))$  ([4], Sect. 8.4.1.4). Figure 7.28 shows with solid and dotted lines the exact and approximate mean and standard deviation of  $X(t)$ ,  $t \in [0, \tau]$  for  $Y \sim U(-0.5, 0.5)$ ,  $Q \sim U(0.5, 1.5)$ ,





**Fig. 7.28** Exact and approximate mean and variance of  $X$  for  $\beta = 1$ ,  $X(0) = 0$ ,  $Y \sim U(-0.5, 0.5)$ ,  $Q \sim U(0.5, 1.5)$ , and  $\tau = 2$

$\tau = 2$ ,  $\beta = 1$ , and  $X(0) = 0$ . The dotted and solid lines coincide at the scale of the figure. The approximate solution consists of the first two terms of the above Neumann series. The approximate mean and standard deviation are accurate in this case.  $\diamond$

*Proof* The integral form of the version  $\dot{X}(t) = -\beta X(t) + [-YX(t) + Q]$  of the defining equation for  $X(t)$  is

$$\begin{aligned}
 X(t) &= X(0)e^{-\beta t} + \int_0^t e^{-\beta(t-s)}(-YX(s) + Q)ds \\
 &= \frac{Q}{\beta}(1 - e^{-\beta t}) - Y e^{-\beta t} \int_0^t e^{\beta s} X(s)ds \\
 &= f(t) + \lambda \int_0^t K(t, s)X(s)ds,
 \end{aligned} \tag{7.90}$$

so that  $X(t)$  satisfies a Fredholm equation. If the kernel  $K$  is square integrable on  $[0, \tau] \times [0, \tau]$  and  $|\lambda| < \|K\|^{-1}$ ,  $\|K\|^2 = \int_{[0, \tau]^2} K(t, s) dt ds$ , then  $X(t)$  admits the infinite series representation in (7.89). Useful information on Neumann series can be found in many text books ([4], Sect. 8.4.1.4, [48], pp. 266–269, [49], Chap. 3, and [50], pp. 49–53).  $\blacktriangle$

## 7.5 Applications

Methods for solving stochastic equations with random coefficients and input developed in the first part of this chapter are applied to study problems involving stochastic stability, noise induced transitions, uncertain dynamic systems, and reliability for degrading systems. Numerical examples are presented to illustrate the implementation of various methods and assess their accuracy.

### 7.5.1 Stochastic Stability

Instability is the cause of spectacular failures in structural and mechanical systems, and can occur under time invariant/variant, deterministic/random actions. A system is said to be stable under a specified action if its response remains bounded at all times. If a real-valued process  $X(t)$  denotes a system response, we say that the system is stable if for arbitrary  $\varepsilon > 0$  there exists  $\delta_\varepsilon > 0$  such that  $|X(t)| \leq \varepsilon$ ,  $t > 0$ , if  $|X(0)| \leq \delta_\varepsilon$ . The system is said to be asymptotically stable if  $|X(t)| \rightarrow 0$  in some sense as  $t \rightarrow \infty$ . Our discussion is limited to the asymptotic stability of the trivial solution  $X(t) = 0$  for dynamic systems driven by white noise. A broad range of results on the stability of dynamic systems under deterministic and stochastic actions can be found in [6, 51–53].

Let  $X(t)$  be a real-valued stochastic process defined by (7.1) with drift and diffusion coefficients depending on only  $X(t)$  rather than  $X(t)$  and  $Y(t)$ . Moreover, the drift and diffusion coefficients are linear in the state, that is,  $X(t)$  is the solution of

$$dX(t) = -\beta X(t-)dt + \sigma_1 X(t-)dB(t) + \sigma_2 X(t-)dC(t), \quad t \geq 0, \quad (7.91)$$

where  $\beta, \sigma_1, \sigma_2 > 0$  are constants,  $X(t-) = \lim_{s \uparrow t} X(s)$  denotes the left limit of the state at time  $t$ ,  $B(t)$  denotes a standard Brownian motion, and

$$C(t) = \sum_{k=1}^{N(t)} Y_k, \quad t \geq 0, \quad (7.92)$$

is a compound Poisson process with iid jumps  $\{Y_k\}$  and jump times  $0 = T_0 < T_1 < \dots < T_k < \dots$  corresponding to a homogeneous Poisson process  $N(t)$  with intensity  $\lambda > 0$ . It is assumed that  $B(t)$  and  $C(t)$  are mutually independent and that initial state  $X(0)$  is independent of these processes.

In random vibration, (7.91) is said to have multiplicative noise since its diffusion coefficients are state dependent. Note also that (7.91) can be viewed as a stochastic differential equation with random coefficients of the type in (7.1) in which the coordinates of  $Y(t)$  are the formal derivatives of  $B(t)$  and  $C(t)$ .

**Theorem 7.21** *If  $1 + \sigma_2 Y_1 > 0$  a.s., the solution of (7.91) is*

$$X(t) = X(0) \exp \left[ -(\beta + \sigma_1^2/2)t + \sigma_1 B(t) + C^*(t) \right], \quad t \geq 0, \quad (7.93)$$

where  $C^*(t) = \sum_{k=1}^{N(t)} Y_k^*$  and  $Y_k^* = \ln(1 + \sigma_2 Y_k)$ .

*Proof* The Itô formula for semimartingales (5.7) applied to

$$V(t) = X(0) \exp[\rho t + \sigma_1 B(t) + C^*(t)], \quad t \geq 0,$$

viewed as a function  $g(t, u, v) = V(t)$  of  $t, u = B(t)$ , and  $v = C^*(t)$  gives

$$\begin{aligned}
V(t) - V(0) = & \int_{0+}^t \frac{\partial g(s, B(s), C^*(s-))}{\partial s} ds + \int_{0+}^t \frac{\partial g(s, B(s), C^*(s-))}{\partial u} dB(s) \\
& + \int_{0+}^t \frac{\partial g(s, B(s), C^*(s-))}{\partial v} dC^*(s) + \frac{1}{2} \int_{0+}^t \frac{\partial^2 g(s, B(s), C^*(s-))}{\partial u^2} d[B, B](s) \\
& + \sum_{0 \leq s < t} \left[ g(s, B(s), C^*(s)) - g(s, B(s), C^*(s-)) - \frac{\partial g(s, B(s), C^*(s-))}{\partial v} \Delta C^*(s) \right],
\end{aligned}$$

where  $\Delta C^*(s) = C^*(s) - C^*(s-)$  denotes the jump of  $C^*$  at  $s$ . This equation simplifies to

$$\begin{aligned}
V(t) - V(0) = & \rho \int_{0+}^t V(s-) ds + \sigma_1 \int_{0+}^t V(s-) dB(s) \\
& + \frac{\sigma_1^2}{2} \int_{0+}^t V(s-) ds + \sum_{0 \leq s < t} [V(s-) (\exp(\Delta C^*(s)) - 1)],
\end{aligned}$$

by using the definition of  $V(t)$ , since  $d[B, B](s) = ds$ ,

$$\begin{aligned}
\int_{0+}^t [\partial g(s, B(s), C^*(s-)) / \partial v] dC^*(s) &= \sum_{0 \leq s < t} [\partial g(s, B(s), C^*(s-)) / \partial v] \Delta C^*(s), \\
g(s, B(s), C^*(s)) - g(s, B(s), C^*(s-)) &= V(s-) [\exp(\Delta C^*(s)) - 1],
\end{aligned}$$

and  $\exp(\Delta C^*(s)) - 1$  is equal to  $\exp(Y_k^*) - 1 = \sigma_2 Y_k$  if  $s = T_k$  and 0 if  $s$  differs from the jump times of  $N(t)$ , that is,  $\exp(\Delta C^*(s)) - 1 = \sigma_2 \Delta C(s)$ . The differential form of above integral equation is

$$dV(t) = (\rho + \sigma_1^2/2) V(t-) dt + \sigma_1 V(t-) dB(t) + \sigma_2 V(t-) dC(t),$$

so that it coincides with defining equation of  $X(t)$  for  $\rho = -\beta - \sigma_1^2/2$ .  $\blacktriangle$

*Example 7.41* The solution of (7.91) with  $\sigma_2 = 0$  in (7.93) can be obtained in an alternative way by using Stratonovich rather than Itô calculus. The Stratonovich version of this equation is

$$dX(t) = (-\beta - \sigma_1^2/2) X(t) dt + \sigma_1 X(t) dB(t), \quad t \geq 0, \quad (7.94)$$

by using the Wong-Zakai correction (Theorem 5.10), which gives (7.93) with  $\sigma_2 = 0$  by following the rules of the classical calculus.  $\diamond$

The diffusion process  $X(t)$  defined by (7.91) with  $\sigma_2 = 0$  is referred to as geometric Brownian motion. By analogy, we call  $X(t)$  in this equation with  $\sigma_1 = 0$  geometric compound Poisson process. In the remainder of this section, we examine the asymptotic stability of the trivial solution of (7.91).

**Theorem 7.22** *Under the assumptions in Theorem 7.21, the trivial solution of (7.91) is asymptotically stable a.s., that is,  $\lim_{t \rightarrow \infty} X(t) = 0$  a.s., if*

$$-(\beta + \sigma_1^2/2) + \lambda E[\ln(1 + \sigma_2 Y_1)] < 0. \quad (7.95)$$

*Proof* Consider the alternative form

$$X(t) = X(0)\exp\left[t\left(-\beta - \sigma_1^2/2 + \sigma_1 B(t)/t + C^*(t)/t\right)\right] = X(0)e^{tZ(t)} \quad (119)$$

of (7.93). We need to establish conditions under which  $Z(t)$  is strictly negative a.s. at large times, and conclude that, under these conditions, the trivial solution of (7.91) is asymptotically stable a.s.

The facts  $B(t)/t \rightarrow 0$  and  $C^*(t)/t \rightarrow \lambda E[\ln(1 + \sigma_2 Y_1)]$  a.s. as  $t \rightarrow \infty$  ([54], p. 496 and Theorem 3.3.2, p. 189) imply  $\lim_{t \rightarrow \infty} Z(t) = -\beta - \sigma_1^2/2 + \lambda E[\ln(1 + \sigma_2 Y_1)]$  a.s., so that  $Z(t) < 0$  at large times if  $-\beta - \sigma_1^2/2 + \lambda E[\ln(1 + \sigma_2 Y_1)] < 0$ , that is, the condition in (7.95). Note that this condition involves both system and noise properties, through the parameter  $\beta$ , the intensity  $\sigma_1$  of  $B(t)$ , the intensity  $\sigma_2$  of  $C(t)$ , and the size and the frequency of the jumps of  $C(t)$ .  $\blacktriangle$

With the notation in the previous theorem,  $|X(t)/X(0)| = \exp(tZ(t))$  so that  $\ln(|X(t)/X(0)|)/t = Z(t)$ . The limit  $\lambda_{LE} = \lim_{t \rightarrow \infty} \ln(|X(t)/X(0)|)/t = \lim_{t \rightarrow \infty} Z(t)$ , referred to as Lyapunov exponent, is equal to the left side of the inequality in (7.95). If  $\lambda_{LE} < 0$ , the trivial solution of (7.91) is asymptotically stable a.s. Lyapunov exponents are used extensively in stochastic stability studies ([6], Chap. 8).

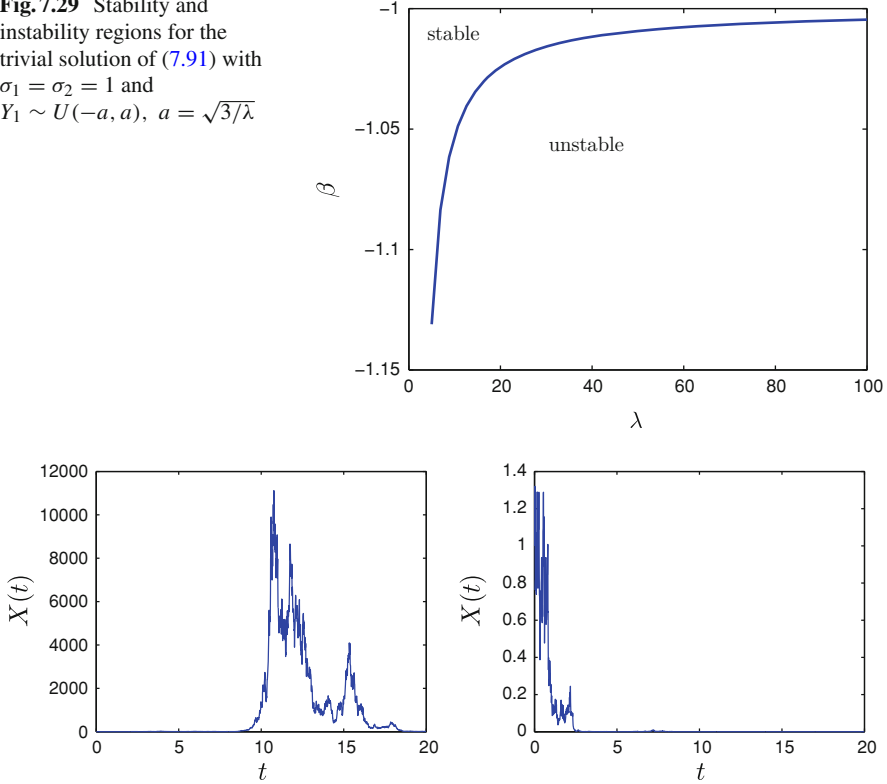
*Example 7.42* For  $\sigma_1 = \sigma_2 = 1$  and  $Y_1 \sim U(-a, a)$  with  $a = \sqrt{3/\lambda}$  such that  $1 + Y_1 > 0$  a.s., the stability condition in (7.93) becomes  $-(\beta + 1/2) + \lambda E[\ln(1 + Y_1)] < 0$  or, equivalently,  $\beta > \beta(\lambda) = -1/2 + \lambda E[\ln(1 + Y_1)]$ . The solid line in Fig. 7.29 partition the  $(\beta, \lambda)$  plane in stability and instability regions for the trivial solution of (7.91).

The stability region for the trivial solution of (7.91) is larger than the stability regions for the trivial solutions of this equation under only Gaussian ( $\sigma_2 = 0$ ) or Poisson ( $\sigma_1 = 0$ ) white noise, that is, the regions  $\beta > -1/2$  and  $\beta > \lambda E[\ln(1 + Y_1)]$ . Figure 7.30 shows samples of  $X(t)$  corresponding to  $(\beta = -0.3, \sigma_1 = 1, \sigma_2 = 0)$  (left panel), and  $(\beta = -0.3, \sigma_1 = 0, \sigma_2 = 1, \lambda = 20)$  (right panel), that is, parameters for which the trivial solutions are stable. The samples in the figure approach 0 as time increases in agreement with the stability condition in (7.95).  $\diamond$

*Example 7.43* Let  $X(t)$  be the solution of (7.91) with  $\sigma_2 = 0$ , so that  $X(t)$  is a geometric Brownian motion. The moment of order  $r \geq 1$  of  $X(t) \mid (X(0) = x)$  is  $E[X(t)^r] = x^r \exp\left[(-\beta r + r(r-1)\sigma_1^2/2)t\right]$  so that it is stable as  $t \rightarrow \infty$  if  $-\beta r + r(r-1)\sigma_1^2/2 < 0$ . Hence, the moments of order  $r = 1, 2$ , and  $3$  are stable if  $\beta > 0$ ,  $\beta > \sigma_1^2/2$ , and  $\beta > \sigma_1^2$ , respectively. Note that  $\beta$  must satisfy distinct conditions for the stability of moments of different orders and that moment stability provides limited if any information on sample stability, that is, the stability condition in (7.95).  $\diamond$

**Theorem 7.23** If  $C(t)$  in (7.91) is replaced by  $C_n(t) = \sum_{k=1}^{N_n(t)} Y_{n,k}$ ,  $n = 1, 2, \dots$ , where  $N_n(t)$  are homogeneous Poisson processes with intensities  $\lambda_n$  increasing

**Fig. 7.29** Stability and instability regions for the trivial solution of (7.91) with  $\sigma_1 = \sigma_2 = 1$  and  $Y_1 \sim U(-a, a)$ ,  $a = \sqrt{3/\lambda}$



**Fig. 7.30** Samples  $X(t)$  for  $(\beta = -0.3, \sigma_1 = 1, \sigma_2 = 0)$  (left panel) and  $(\beta = -0.3, \sigma_1 = 0, \sigma_2 = 1, \lambda = 20)$  (right panel)

monotonically with  $n$  such that  $\lim_{n \rightarrow \infty} \lambda_n = \infty$ , and  $Y_{n,k}$  are iid random variables uniformly distributed in  $[-a_n, a_n]$ ,  $a_n = \sqrt{3/\lambda_n}$ , then  $C_n(t)$  converges weakly to a standard Brownian motion as  $n \rightarrow \infty$ . The stability condition in (7.95) becomes  $\beta > -(\sigma_1^2 + \sigma_2^2)/2$  as  $n \rightarrow \infty$ .

*Proof* Note that the processes  $C_n(t)$  and  $B(t)$  are equal in the second moments sense since  $Y_{n,1} \sim U(-\sqrt{3/\lambda_n}, \sqrt{3/\lambda_n})$  by assumption, so that the mean and covariance functions of  $C_n(t)$  are  $E[C_n(t)] = 0$  and  $E[C_n(s)C_n(t)] = \lambda_n(s \wedge t)E[Y_{n,1}^2] = s \wedge t$ .

We first show that, asymptotically as  $n \rightarrow \infty$ ,  $C_n(t)$  becomes a version of a standard Brownian motion  $B(t)$ . Since  $C_n(t)$  and  $B(t)$  have stationary independent increments, it is sufficient to show that their marginal characteristic functions coincides as  $n \rightarrow \infty$ , that is,  $\lim_{n \rightarrow \infty} \varphi_{C_n(t)}(u) = \varphi_{B(t)}(u)$ , where  $\varphi_{C_n(t)}(u) = \exp[\lambda_n t (E[e^{iuY_{n,1}}] - 1)]$  and  $\varphi_{B(t)}(u) = \exp(-u^2 t/2)$ . Since exponential are continuous function, it is sufficient to show  $\lim_{n \rightarrow \infty} \lambda_n (E[e^{iuY_{n,1}}] - 1) = -u^2/2$ . The expectation of the Taylor expansion of  $\exp(iuY_{n,1})$  about  $u = 0$  is  $E[e^{iuY_{n,1}}] =$

$1 - \frac{u^2}{2} E[Y_{n,1}^2] + E[ri u Y_{n,1}]$ , where  $|r(v)| \leq v^2 \alpha(|v|)$  and  $\alpha : [0, \infty) \rightarrow \mathbb{R}$  is an increasing function such that  $\lim_{v \downarrow 0} \alpha(v) = 0$ . We have

$$\lambda_n (E[e^{iu Y_{n,1}}] - 1) = \lambda_n \left( -\frac{u^2}{2} E[Y_{n,1}^2] + E[ri u Y_{n,1}] \right) = -\frac{u^2}{2} + \lambda_n E[ri u Y_{n,1}].$$

Since  $|\lambda_n E[ri u Y_{n,1}]| \leq \lambda_n E[|(iu Y_{n,1})^2 \alpha(|iu Y_{n,1}|)|] \leq \lambda_n u^2 E[Y_{n,1}^2] \alpha(u \sqrt{3/\lambda_n})$  and  $\alpha(u \sqrt{3/\lambda_n}) \rightarrow 0$  as  $n \rightarrow \infty$ , we have  $\lim_{n \rightarrow \infty} \varphi_{C_n(t)}(u) = \varphi_{B(t)}(u)$ . It is possible to extend the proof to distributions other than  $Y_{n,1} \sim U(-\sqrt{3/\lambda_n}, \sqrt{3/\lambda_n})$ .

We now show that the sequence of compound Poisson processes  $\{C_n(t)\}$  converges weakly to a standard Brownian motion  $B(t)$  in  $[0, 1]$ . The consideration of time interval  $[0, 1]$  is not restrictive. Let  $C[0, 1]$  be the space of real-valued continuous function defined in  $[0, 1]$ . This space with the uniform metric  $\rho(x, y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|$ ,  $x, y \in C[0, 1]$ , is a separable complete metric space ([55], Appendix 1). Denote by  $\mathcal{C}$  the Borel  $\sigma$ -field generated by the family of open balls  $\{y \in C[0, 1] : \rho(x, y) < r\}$ ,  $r > 0$ , centered at  $x \in C[0, 1]$ , that is, the  $\sigma$ -field generated by the uniform topology.

Let  $D[0, 1]$  be the space of real-valued right continuous with left limit functions defined in  $[0, 1]$ , so that  $x(t+) = \lim_{s \downarrow t} x(s) = x(t)$  and  $x(t-) = \lim_{s \uparrow t} x(s)$  exists at all times for all  $x \in D[0, 1]$ . It is assumed that  $x(0) = x(0+)$  for all members of  $D[0, 1]$ . Let  $\mathcal{E}$  be the set of continuous, strictly increasing functions  $\xi : [0, 1] \rightarrow [0, 1]$  such that  $\xi(0) = 0$  and  $\xi(1) = 1$ . The function  $\rho_D(x, y) = \inf_{\xi \in \mathcal{E}} |\sup_{0 \leq t \leq 1} |x(t) - y(t)| + \sup_{0 \leq t \leq 1} |t - \xi(t)|$  is a metric on  $D[0, 1]$ . Let  $\mathcal{D}$  be the Borel  $\sigma$ -field generated by the open balls  $\{y \in D[0, 1] : \rho_D(x, y) < r\}$ ,  $r > 0$ , centered at  $x \in D[0, 1]$ , that is, the  $\sigma$ -field generated by the Skorokhod topology. Note that (1) a sequence  $\{x_n \in D[0, 1], n = 1, 2, \dots\}$  converges to an element  $x \in D[0, 1]$  if and only if there exists functions  $\xi_n \in \mathcal{E}$  such that  $\lim_{n \rightarrow \infty} x_n(\xi_n(t)) = x(t)$  and  $\lim_{n \rightarrow \infty} \xi_n(t) = t$  both uniformly in  $t$ , (2) the Skorokhod convergence is weaker than the uniform convergence, and (3) Skorokhod topology relativized to  $C[0, 1]$  coincides with the uniform topology.

Since the samples of  $B$  and  $C_n$  are elements of  $C[0, 1]$  and  $D[0, 1]$  and  $C[0, 1] \subset D[0, 1]$ , we need to prove the convergence of  $C_n$  to  $B$  in  $D[0, 1]$ . The measure  $W$  of  $B$  on  $(C[0, 1], \mathcal{C})$  can be extended to  $(D[0, 1], \mathcal{D})$  by setting  $W(A) = W(A \cap C[0, 1])$  for  $A \in \mathcal{D}$  ([55], Sect. 16). The weak convergence  $C_n \Rightarrow B$  follows from Theorem 15.1 in [55], stating that, if a family of probability measures  $\{P_n\}$  is tight and if  $P_n \pi_{t_1 \dots t_k}^{-1} \Rightarrow P \pi_{t_1 \dots t_k}^{-1}$  holds whenever  $t_1, \dots, t_k$  all lie in  $T_P$ , then  $P_n \Rightarrow P$ . The set  $T_P$  consists of those times  $t \in [0, 1]$  at which the projection function  $\pi_t$  is continuous except at points forming a set of  $P$ -measure 0.

The convergence  $P_n \pi_{t_1 \dots t_k}^{-1} \Rightarrow P \pi_{t_1 \dots t_k}^{-1}$  of the finite dimensional distributions of  $C_n$  to those of  $B$  corresponding to arbitrary  $t_1, \dots, t_k$  has already been shown. It remains to show that the sequence of processes  $\{C_n\}$  is tight. The sequence  $\{C_n\}$  is tight if (1) for each positive  $\eta$ , there exists  $a$  such that  $P(\sup_{0 \leq t \leq 1} |C_n(t)| > a) \leq \eta$ ,  $n \geq 1$ , and (2) for each positive  $\varepsilon$  and  $\eta$ , there exists  $\delta \in (0, 1)$  and  $n_0 \in \{1, 2, \dots\}$  such that  $P(W'_n(\delta) \geq \varepsilon) \leq \eta$ ,  $n \geq n_0$ , where

$$W_n(\delta) = \sup\{|C_n(s') - C_n(s'')| : s', s'' \in [t, t + \delta]\},$$

$$W'_n(\delta) = \inf_{\{t_i\}} \max_{0 < i \leq r} W_n[t_{i-1}, t_i),$$

$0 \leq t_0 < t_1 < \dots < t_r = 1$  is such that  $t_i - t_{i-1} > \delta$ , and  $x \in D[0, 1]$ . For  $a > 0$  arbitrary, we have

$$\begin{aligned} P\left(\sup_{0 < s \leq t} |C_n(s)| \geq a\right) &\leq P\left(\sup_{0 < s \leq t} C_n(s) \geq a\right) + P\left(\inf_{0 < s \leq t} C_n(s) \leq -a\right) \\ &= 2P\left(\sup_{0 < s \leq t} C_n(s) \geq a\right), \end{aligned}$$

by the postulated symmetry of the density of  $C_n$ . Let  $T_n(a) = \inf\{t \geq 0 : C_n(t) \geq a\}$  be the first time when  $C_n$  exceeds  $a$  and note that

$$P(C_n(t) \geq a) = P(C_n(t) \geq a \mid T_n(a) \leq t)P(T_n(a) \leq t)$$

since  $P(C_n(t) \geq a \mid T_n(a) > t) = 0$ . The inequality  $C_n(T_n(a)) \geq a$  and the symmetry of the density of  $C_n$  imply  $P(C_n(t) \geq a \mid T_n(a) \leq t) \geq 1/2$ , so that

$$\begin{aligned} P\left(\sup_{0 < s \leq t} C_n(s) \geq a\right) &= P(T_n(a) \leq t) \leq 2P(C_n(t) \geq a), \\ P\left(\sup_{0 < s \leq t} |C_n(s)| \geq a\right) &\leq 4P(C_n(t) \geq a) = 2P(|C_n(t)| \geq a), \end{aligned}$$

and

$$P\left(\sup_{0 < s \leq t} |C_n(s)| \geq a\right) \leq 2P(|C_n(t)| \geq a) \leq 2 \frac{E[C_n(t)^2]}{a^2} = \frac{2\lambda_n t E[Y_{n,1}^2]}{a^2} = \frac{2t}{a^2}$$

by Chebyshev's inequality and properties of  $C_n(t)$ . Hence, for each  $\eta > 0$  there exists  $a = (2t/\eta)^{1/2}$  satisfying the first condition.

For the second condition, note that  $W_n[T_{n,k-1}, T_{n,k}) = 0$  in the time intervals between consecutive jumps of  $C_n(t)$ , so that  $W'_n(\delta) = \max_{1 \leq k \leq N_n(t)} \{|Y_{n,k}|\}$ . The distribution of  $W'_n(\delta)$  is  $P(W'_n(\delta) \leq y) = \sum_{q=0}^{\infty} (2F_n(y) - 1)^q P(N(t) = q)$ , so that  $P(W'_n(\delta) > \varepsilon) = 1 - \exp[-2\lambda_n t (1 - F_n(\varepsilon))]$ ,  $\varepsilon > 0$ , where  $F_n$  denotes the distribution of  $Y_{n,1}$ . The required condition is satisfied for  $\eta = 1 - \exp[-2\lambda_n t (1 - F_n(\varepsilon))]$  since  $\lambda_n (1 - F_n(\varepsilon)) \rightarrow 0$  as  $n \rightarrow \infty$ . Hence, the sequence of processes  $\{C_n\}$  is tight.

Since the finite dimensional distributions of  $\{C_n\}$  converge to those of  $B$  and the sequence of random elements  $\{C_n\}$  is tight, we conclude that  $\{C_n\}$  converges weakly to  $B$ .  $\blacktriangle$

According to Theorem 7.23, the stability of the trivial solution of (7.91) as  $\lambda \rightarrow \infty$  coincides with that for the trivial solution of  $dX(t) = -\beta X(t)dt + \sigma_1 X(t)dB(t) + \sigma_2 X(t)dB_1(t)$ , where  $B(t)$  and  $B_1(t)$  are independent standard Brownian motions. The trivial solution of the latter equation is stable if  $\beta > -(\sigma_1^2 + \sigma_2^2)/2$ , that is,  $\beta > -1$  for  $\sigma_1 = \sigma_2 = 1$ , a result that is in agreement with the plot in Fig. 7.29, that seems to approach  $-1$  as  $\lambda$  increases.

Let  $X(t)$  be the solution of (7.91) in which the driving noise processes  $B(t)$  and  $C(t)$  are replaced by an  $\alpha$ -stable Lévy process  $L_\alpha(t)$ , that is,

$$dX(t) = -\beta X(t-) dt + \sigma X(t-) dL_\alpha(t), \quad t \geq 0, \quad (7.96)$$

where  $\beta, \sigma > 0$  are constants. An approximate solution of this equation can be obtained by replacing  $L_\alpha(t)$  with

$$L_{\alpha,a}(t) = \sigma(\alpha, a)B(t) + C_{\alpha,a}(t), \quad t \geq 0, \quad (7.97)$$

based on the fact that any  $\alpha$ -stable process can be viewed as a sum of two independent processes, a scaled Brownian motion and a compound Poisson process  $C_{\alpha,a}$  corresponding to the jumps of  $L_{\alpha,a}$  with magnitude larger than an arbitrary constant  $a > 0$ . We have seen in Sect. 3.7.6.3 that the accuracy of the representation in (7.97) is remarkable. The differential equation for  $X(t)$  in (7.96) with  $L_{\alpha,a}(t)$  in place of  $L_\alpha(t)$  becomes

$$dX(t) = -\beta X(t-) dt + \sigma X(t-) (\sigma(\alpha, a) dB(t) + dC_{\alpha,a}(t)), \quad t \geq 0, \quad (7.98)$$

that is, a stochastic differential equation of the type in (7.91). Theorem 7.22 cannot be applied to study the stability of the trivial solution of this equation since the jumps of  $C_{\alpha,a}(t)$  do not satisfy the conditions of this theorem.

*Example 7.44* Let  $X(t)$  be the solution of  $dX(t) = -\beta X(t-) dt + \sigma X(t-) dL_\alpha(t)$ ,  $t \geq 0$ , where  $\beta, \sigma > 0$  are constants and  $L_\alpha(t)$  is an  $\alpha$ -stable process whose increments  $L_\alpha(t) - L_\alpha(s) \sim S_\alpha((t-s)^{1/\alpha}, 0, 0)$ ,  $t > s$ , are  $\alpha$ -stable random variables with scale  $(t-s)^{1/\alpha}$ , skewness 0, and location 0. If  $\alpha \in (1, 2]$ , the mean of  $L_\alpha(t) - L_\alpha(s)$  exists and is 0 ([56], Property 1.2.19), so that the expectation of the defining equation for  $X(t)$  gives  $\dot{\mu}(t) = -\beta\mu(t)$  with the notation  $\mu(t) = E[X(t)]$  since  $X(t-)$  and  $dL_\alpha(t)$  are independent and  $E[dL_\alpha(t)] = 0$ . This gives  $\mu(t) = \mu(0)\exp(-\beta t)$  so that  $\mu(t) \rightarrow 0$  as  $t \rightarrow \infty$ , that is, the first moment of  $X(t)$  is stable.  $\diamond$

*Example 7.45* Suppose  $X(t)$  is a real-valued stochastic process defined by

$$\ddot{X}(t) + 2\beta\dot{X}(t) + [1 + \sigma_1 W_1(t)]X(t) = \sigma_2 W_2(t), \quad t \geq 0, \quad (7.99)$$

where  $\beta > 0$ ,  $\sigma_1$ , and  $\sigma_2$  are real constants, and  $W_1$  and  $W_2$  denote independent Gaussian white noise processes. The moments  $\mu(p, q; t) = E[X_1(t)^p X_2(t)^q]$  of the bivariate process  $Z(t) = (X_1(t) = X(t), X_2(t) = \dot{X}(t))$  satisfy the differential equations

$$\begin{aligned} \dot{\mu}(p, q; t) = & p\mu(p-1, q+1; t) - q\mu(p+1, q-1; t) - 2q\beta\mu(p, q; t) \\ & + \frac{q(q-1)\sigma_1^2}{2}\mu(p+2, q-2; t) + \frac{q(q-1)\sigma_2^2}{2}\mu(p, q-2; t) \end{aligned}$$

with the convention  $\mu(p, q; t) = 0$  if  $p < 0$  and/or  $q < 0$ . This is a closed system of equations, so that moments of any order of  $X(t)$  and  $\dot{X}(t)$  can be calculated exactly. For example, the moments of order  $p+q=1$  and  $p+q=2$  result from



$$\begin{aligned}\dot{\mu}(1, 0; t) &= \mu(0, 1; t), \\ \dot{\mu}(0, 1; t) &= -\mu(1, 0; t) - 2\beta\mu(0, 1; t)\end{aligned}$$

and

$$\begin{aligned}\dot{\mu}(2, 0; t) &= 2\mu(1, 1; t), \\ \dot{\mu}(1, 1; t) &= \mu(0, 2; t) - \mu(2, 0; t) - 2\beta\mu(1, 1; t), \\ \dot{\mu}(0, 2; t) &= -2\mu(1, 1; t) - 4\beta\mu(0, 2; t) + \sigma_1^2\mu(2, 0; t) + \sigma_2^2.\end{aligned}$$

Differential equations of the type in (7.99) provide useful models in applications. For example, the deflection  $V(x, t)$  of a simply supported beam with constant stiffness  $\chi$  and length  $l > 0$  subjected to a fluctuating axial force  $W(t)$  applied at its ends satisfies the differential equation

$$\chi \frac{\partial^4 V(x, t)}{\partial x^4} = -W(t) \frac{\partial^2 V(x, t)}{\partial x^2} - m \frac{\partial^2 V(x, t)}{\partial t^2},$$

where  $m$  is the beam mass per unit length. For the first buckling mode we have  $V(x, t) = X(t) \sin(\pi x/l)$  so that

$$\ddot{X}(t) + v^2 (1 - W(t)/p_{\text{cr}}) X(t) = 0,$$

where  $v^2 = \pi^4 \chi / (ml^4)$  and  $p_{\text{cr}} = \pi^2 \chi / l^2$  is the first buckling load for the beam.  $\diamond$

*Proof* Since  $W_1(t)$  and  $W_2(t)$  are formal derivatives of standard Brownian processes  $B_1(t)$  and  $B_2(t)$ ,  $Z(t)$  is a diffusion process defined by the stochastic differential equation

$$\begin{cases} dX_1(t) = X_2(t) dt, \\ dX_2(t) = -[X_1(t) + 2\beta X_2(t)] dt - \sigma_1 X_1(t) dB_1(t) + \sigma_2 dB_2(t). \end{cases}$$

The moment equations result from Itô's formula applied to the mapping  $(X_1 X_2) \mapsto X_1^p X_2^q$  by averaging.

Since moment equations for fixed  $p + q$  have the form  $\dot{m}(t) = am(t)$ , the moments of order  $p + q$  of  $Z(t)$  become time invariant as  $t \rightarrow \infty$  if the eigenvalues of  $a$  have negative real parts. For example, the eigenvalue of  $a$  for moments of order  $p + q = 1$  are  $-\beta \pm \sqrt{\beta^2 - 1}$  so that, if  $\beta > 0$ ,  $X(t)$  and  $\dot{X}(t)$  are asymptotically stable in the mean.  $\blacktriangle$

## 7.5.2 Noise Induced Transitions

Let  $x(t)$ ,  $t \geq 0$ , denote the average population per unit area or any other scalar measure for the size of a biological population. According to the Verhulst model,  $x(t)$  satisfies the differential equation

$$\dot{x}(t) = \rho x(t) - x(t)^2, \quad t > 0, \quad (7.100)$$

where  $\rho x(t)$ ,  $\rho \in \mathbb{R}$ , is the rate of population increase/decrease and  $-x(t)^2$  relates to available resources. Positive and negative values of  $\rho$  correspond to favorable and hostile environments. The solution,

$$x(t) = x(0)e^{\rho t} [1 + x(0)(e^{\rho t} - 1)/\rho]^{-1}, \quad (7.101)$$

of (7.100), is asymptotically stable for  $\rho \neq 0$  since  $x_s(t) = \lim_{t \rightarrow \infty} x(t) = 0$  for  $\rho < 0$  and  $x_s(t) = \rho$  for  $\rho > 0$ . The system undergoes a phase transition at  $\rho = 0$  from  $x_s(t) = 0$  for  $\rho < 0$  to  $x_s(t) = \rho$  for  $\rho > 0$ .

Consider an extended version of the Verhulst model in which  $\rho$  is replaced with  $\rho + \sigma dB(t)/dt + \gamma dC(t)/dt$ , where  $\sigma, \gamma$  are constants,  $B(t)$  is a standard Brownian motion, and  $C(t)$  is a compound Poisson process defined by (7.92) with jumps  $Y_k$  that have mean 0 and finite variance. Note that the expectation of the random environment described by  $\rho + \sigma dB(t)/dt + \gamma dC(t)/dt$  is equal to  $\rho$ . For this environment, the population size  $X(t)$  satisfies the stochastic differential equation

$$dX(t) = (\rho X(t-) - X(t-)^2)dt + X(t-)(\sigma dB(t) + \gamma dC(t)), \quad t \geq 0. \quad (7.102)$$

**Theorem 7.24** *The characteristic function  $\varphi(u; t) = E[\exp(iuX(t))]$  of  $X(t)$  in (7.102) is the solution of the integro-differential equation*

$$\begin{aligned} \frac{\partial \varphi(u; t)}{\partial t} = & \rho u \frac{\partial \varphi(u; t)}{\partial u} + iu \frac{\partial^2 \varphi(u; t)}{\partial u^2} + \frac{\sigma^2 u^2}{2} \frac{\partial^2 \varphi(u; t)}{\partial u^2} \\ & + \lambda \left[ \int_{\mathbb{R}} \varphi(u(1 + \gamma y); t) dF(y) - \varphi(u; t) \right], \end{aligned} \quad (7.103)$$

where  $F(y)$  denotes the distribution of the jumps of  $C(t)$ .

*Proof* Arguments similar to those in [4] (Sect. 9.4.3) are used to establish (7.103). Itô's formula applied to the mapping  $X(t) \mapsto \exp(iuX(t))$  gives

$$\begin{aligned} e^{iuX(t)} - e^{iuX(0)} = & \int_{0+}^t iue^{iuX(s-)} dX(s) + \frac{1}{2} \int_{0+}^t (iu)^2 e^{iuX(s-)} d[X, X]^c(s) \\ & + \sum_{0 < s \leq t} \left[ e^{iuX(s)} - e^{iuX(s-)} - iue^{iuX(s-)} \Delta X(s) \right] \\ = & \int_{0+}^t iue^{iuX(s-)} \left( (\rho X(s-) - X(s-)^2) ds + \sigma X(s-) dB(s) \right) \\ & + \frac{1}{2} \int_{0+}^t (iu)^2 e^{iuX(s-)} \sigma^2 X(s-)^2 ds + \sum_{0 < s \leq t} [e^{iuX(s)} - e^{iuX(s-)}] \end{aligned}$$

since the increment of the continuous part of the quadratic variation of  $X(s)$  is  $d[X, X]^c(s) = \sigma^2 X(s-)^2 ds$  and

$$\int_{0+}^t (iu)e^{iuX(s-)} \gamma dC(s) = \sum_{0 < s \leq t} iue^{iuX(s-)} \Delta X(s).$$

The expectation of the above equation can be calculated term by term and gives

$$\begin{aligned} \varphi(u; t) - \varphi(u; 0) &= iuE \left[ \int_{0+}^t (iu e^{iuX(s-)} (\rho X(s-) - X(s-)^2) ds \right] \\ &\quad - \frac{\sigma^2 u^2}{2} E \left[ \int_{0+}^t e^{iuX(s-)} X(s-)^2 ds \right] + E \left[ \sum_{0 < s \leq t} \left( e^{iuX(s)} - e^{iuX(s-)} \right) \right]. \end{aligned}$$

The first three terms on the right side of (7.103) result from the first two expectations in the previous formula by applying Fubini's theorem and using properties of the characteristic function. The fourth term on the right side of (7.103) follows from

$$E \left[ \sum_{0 < s \leq t} \left( e^{iuX(s)} - e^{iuX(s-)} \right) \right] = E \left\{ E \left[ \sum_{k=1}^{N(t)} \left( e^{iuX(T_k)} - e^{iuX(T_k-)} \right) \mid N(t) \right] \right\},$$

$X(T_k) = X(T_k-)(1 + \gamma Y_k)$ , the independence of  $Y_k$  from  $X(T_k-)$ , the property  $P(N(t + \Delta t) - N(t) \geq 1) \simeq \lambda \Delta t$  of  $N(t)$  that holds for  $\lambda \Delta t \ll 1$ , and the fact that, conditional on  $N(t)$ , the jump times  $T_k$  are independent random variables distributed uniformly in  $(0, t)$ . ▲

**Example 7.46** Consider the special case of (7.102) with  $\gamma = 0$ , so that  $X(t)$  is a diffusion process defined by  $dX(t) = (\rho X(t) - X(t)^2)dt + \sigma X(t)dB(t)$ . Then (7.103) becomes a partial differential equations for  $\varphi(u; t)$ . Its Fourier transform (Theorem 7.4) constitutes the Fokker–Planck equation for the density  $f(x; t)$  of  $X(t) \mid (X(0) = x)$ , and admits the stationary solutions  $f_s(x) = \beta x^{2(\rho/\sigma^2-1)} \exp(-2x/\sigma^2)$ ,  $x > 0$ , for  $\rho > \sigma^2/2$  and  $f_s(x) = \delta(x)$  for  $\rho < \sigma^2/2$ , where  $\beta > 0$  is a constant. Figure 7.31 shows densities  $f_s(x)$  for  $\sigma = 1$ ,  $\rho < \sigma^2/2$ ,  $\sigma^2/2 < \rho < \sigma^2$ , and  $\rho > \sigma^2$ . The stationary density changes qualitatively at  $\rho = \sigma^2/2$  and  $\rho = \sigma^2$ . Since similar changes in  $f_s(x)$  result if  $\rho$  is kept constant and noise intensity  $\sigma$  is varied, the qualitative changes in  $f_s(x)$  are referred to as noise induced transitions. ◇

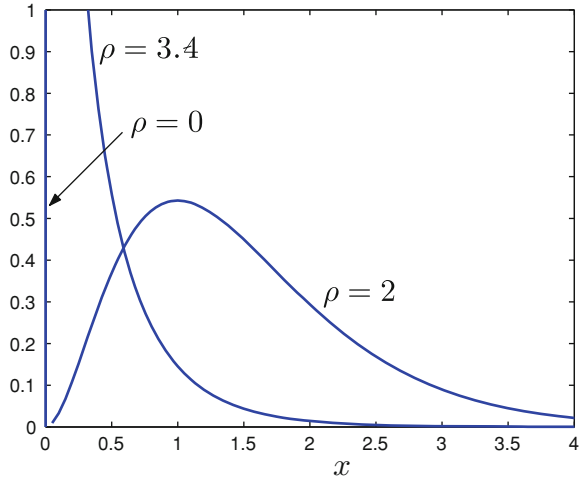
If  $\sigma = 0$ , (7.103) takes the form

$$\frac{\partial \varphi(u; t)}{\partial t} = \rho u \frac{\partial \varphi(u; t)}{\partial u} + iu \frac{\partial^2 \varphi(u; t)}{\partial u^2} + \lambda \left[ \int_{\mathbb{R}} \varphi(u(1 + \gamma y); t) dF(y) - \varphi(u; t) \right], \quad (7.104)$$

and cannot be solved analytically. The integro-differential equation (7.104) can be used to find  $X(t)$  for the Gaussian random environment  $\rho + \gamma dB(t)/dt$  by increasing indefinitely the arrival rate of the jumps of  $C(t)$  and simultaneously reducing their size, as shown by the following theorem.

**Theorem 7.25** *If  $E[Y_1] = 0$ ,  $\lambda E[Y_1^2] = 1$ ,  $P(|Y_1| \leq a) = 1$ , and  $a > 0$  is such that  $a \rightarrow 0$  as  $\lambda \rightarrow \infty$ , then (7.104) becomes*

**Fig. 7.31** Stationary density  $f_s(x)$  for  $\sigma = 1$ ,  $\rho = 0$  ( $\rho < \sigma^2/2$ ),  $\rho = 3/4$  ( $\sigma^2/2 < \rho < \sigma^2$ ), and  $\rho = 2$  ( $\rho > \sigma^2$ )



$$\frac{\partial \varphi(u; t)}{\partial t} = \rho u \frac{\partial \varphi(u; t)}{\partial u} + iu \frac{\partial^2 \varphi(u; t)}{\partial u^2} + \frac{\gamma^2 u^2}{2} \frac{\partial^2 \varphi(u; t)}{\partial u^2}, \quad (7.105)$$

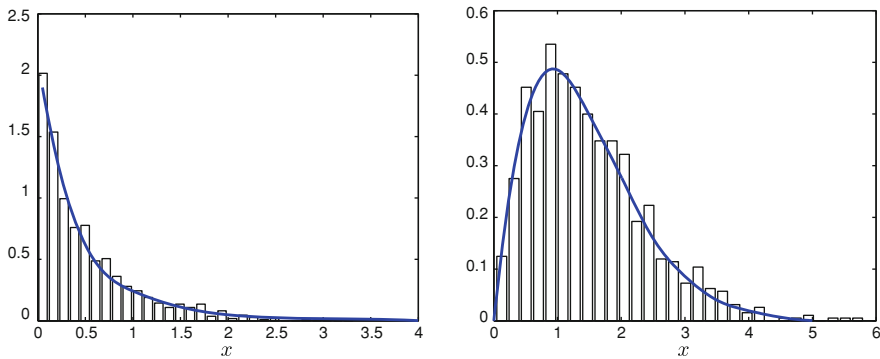
as  $\lambda \rightarrow \infty$ , that is,  $X(t)$  is the solution of the Verhulst model with Gaussian rate  $\rho + \gamma dB(t)/dt$ .

*Proof* The jumps  $Y_k \sim U(-a, a)$ ,  $a = \sqrt{3/\lambda}$ , have the required properties. For these jumps, the term in square brackets of (7.104) is

$$\begin{aligned} & \lambda \left[ \int_{\mathbb{R}} \varphi(u(1 + \gamma y); t) dF(y) - \varphi(u; t) \right] \\ &= \lambda \left[ \int_{\mathbb{R}} \left( \varphi(u; t) + (u\gamma y) \frac{\partial \varphi(u; t)}{\partial u} + \frac{(u\gamma y)^2}{2} \frac{\partial^2 \varphi(u; t)}{\partial u^2} + r(u\gamma y) \right) dF(y) - \varphi(u; t) \right] \\ &= \lambda \left[ u\gamma \frac{\partial \varphi(u; t)}{\partial u} E[Y_1] + \frac{(u\gamma)^2}{2} \frac{\partial^2 \varphi(u; t)}{\partial u^2} E[Y_1^2] + \int_{\mathbb{R}} r(u\gamma y) dF(y) \right] \\ &= \frac{(u\gamma)^2}{2} \frac{\partial^2 \varphi(u; t)}{\partial u^2} + \lambda \int_{\mathbb{R}} r(u\gamma y) dF(y), \end{aligned}$$

where  $\varphi(u(1 + \gamma y); t)$  has been expanded in Taylor's series about  $u$ ,  $|r(u\gamma y)| \leq (u\gamma y)^2 \alpha(|u\gamma y|)$ , and  $\alpha: [0, \infty) \rightarrow [0, \infty)$  is a monotonically increasing function such that  $\lim_{\xi \downarrow 0} \alpha(\xi) = 0$ . Since

$$\begin{aligned} \left| \lambda \int_{\mathbb{R}} r(u\gamma y) dF(y) \right| &\leq \lambda \int_{\mathbb{R}} |r(u\gamma y)| dF(y) \leq \lambda \int_{\mathbb{R}} (u\gamma y)^2 \alpha(|\gamma u|) dF(y) \\ &\leq \lambda u^2 \gamma^2 \alpha(a) \int_{\mathbb{R}} y^2 dF(y) = u^2 \gamma^2 \alpha(a) \rightarrow 0, \quad \text{as } a \rightarrow 0 \ (\lambda \rightarrow \infty), \end{aligned}$$



**Fig. 7.32** Histograms and marginal densities of  $X(t)$  at  $t = 10$  for  $\lambda = 12$ ,  $\rho = 1$ ,  $\sigma = 0$ , and  $\gamma = 1$  (left panel) and  $\lambda = 12$ ,  $\rho = 2$ ,  $\sigma = 0$ , and  $\gamma = 1$  (right panel)

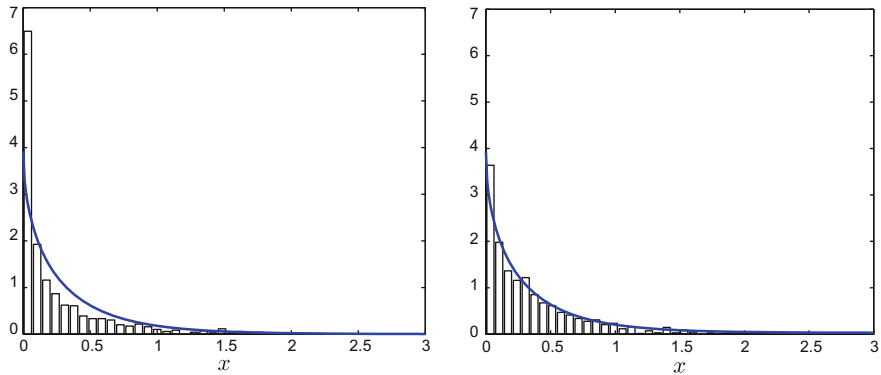
we have

$$\lim_{\lambda \rightarrow \infty} \lambda \left[ \int_{\mathbb{R}} \varphi(u(1 + \gamma y); t) dF(y) - \varphi(u; t) \right] = \frac{(u\gamma)^2}{2} \frac{\partial^2 \varphi(u; t)}{\partial u^2},$$

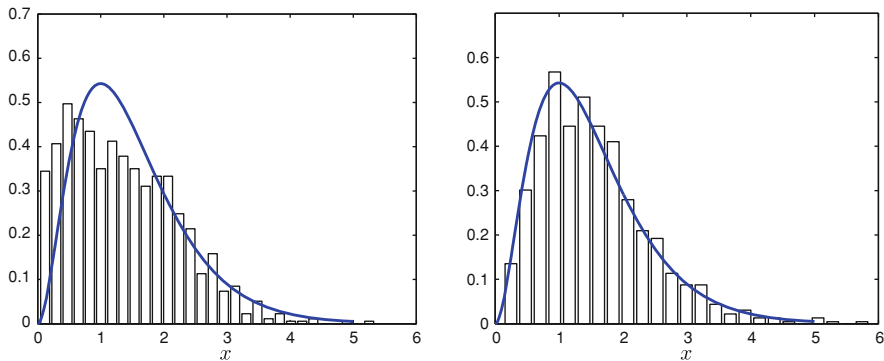
which yields (7.105). ▲

*Example 7.47* Let  $X(t)$  be the solution of (7.102) with  $\sigma = 0$ , so that it satisfies the stochastic differential equation  $dX(t) = (\rho X(t-) - X(t-)^2)dt + \gamma X(t-)dC(t)$ . The heavy solid lines in Fig. 7.32 are Fourier transforms  $f_s(x)$  of stationary solutions  $\varphi_s(u) = \lim_{t \rightarrow \infty} \varphi(u; t)$  of (7.104) for  $\lambda = 12$ ,  $\rho = 1$ ,  $\sigma = 0$ , and  $\gamma = 1$  (left panel), and  $\lambda = 12$ ,  $\rho = 2$ ,  $\sigma = 0$ , and  $\gamma = 1$  (right panel). The stationary characteristic functions  $\varphi_s(u)$  are numerical solutions of (7.104) for these parameter values. The histograms have been constructed from 1000 independent samples of  $X(t)$  at  $t = 10$  defined by (7.102) with  $\sigma = 0$  and  $\gamma = 1$ . They follow closely the densities  $f_s(x)$  and capture the phase transition phenomenon exhibited by  $f_s(x)$ . The samples of  $X(t)$  have been generated by the fixed time step integration scheme in [57] using 1000 equal time steps in  $[0, 10]$ . Time  $t = 10$  has been selected to assure that  $X(t)$  has essentially reached stationarity.

The heavy solid lines in Figs. 7.33 and 7.34 are the stationary densities of  $X(t)$  in (7.102) for  $\rho = 3/4$  and  $\rho = 2$ , respectively. Both densities are for  $\sigma = 1$  and  $\gamma = 0$ , so that they correspond to  $X(t)$  under Gaussian white noise, and have the expression  $f_s(x) = \beta x^{2(\rho/\sigma^2 - 1)} \exp(-2x/\sigma^2)$ ,  $x > 0$ , since  $\rho > \sigma^2/2$ . The histograms have been obtained from 1000 independent samples of  $X(t)$  at  $t = 10$  defined by (7.102) with  $\sigma = 0$  and  $\gamma = 1$ , that is,  $X(t)$  under Poisson white noise. The histograms in Fig. 7.33 are for  $\lambda = 5$  and  $\rho = 3/4$  (left panel) and  $\lambda = 30$  and  $\rho = 3/4$  (right panel). The histograms in Fig. 7.34 are for  $\lambda = 5$  and  $\rho = 2$  (left panel) and  $\lambda = 30$  and  $\rho = 2$  (right panel). The fixed time step scheme in [57] with 1000 steps in  $[0, 10]$  has been used to generate samples of  $X(t)$ . The histograms are consistent with



**Fig. 7.33** Histograms of  $X(t)$  at  $t = 10$  for  $\lambda = 5$ ,  $\rho = 3/4$ ,  $\sigma = 0$ , and  $\gamma = 1$  (left panel), and  $\lambda = 30$ ,  $\rho = 3/4$ ,  $\sigma = 0$ , and  $\gamma = 1$  (right panel). Solid lines in both panels are the stationary density of  $X(t)$  for  $\rho = 3/4$ ,  $\sigma = 1$ , and  $\gamma = 0$



**Fig. 7.34** Histograms of  $X(t)$  at  $t = 10$  for  $\lambda = 5$ ,  $\rho = 2$ ,  $\sigma = 0$ , and  $\gamma = 1$  (left panel), and  $\lambda = 30$ ,  $\rho = 2$ ,  $\sigma = 0$ , and  $\gamma = 1$  (right panel). Solid lines in both panels are the stationary density of  $X(t)$  for  $\rho = 2$ ,  $\sigma = 1$ , and  $\gamma = 0$

Theorem 7.25 in the sense that they differ from  $f_s(x)$  for  $\lambda = 5$  but follow closely this density for  $\lambda = 30$ .  $\diamond$

**Theorem 7.26** If  $\lim_{|u| \rightarrow \infty} u\varphi_s(u) = 0$ ,  $\lim_{|u| \rightarrow \infty} u\varphi'_s(u) = 0$ , and (7.104) admits a unique stationary solution  $\varphi_s(u) = \lim_{t \rightarrow \infty} \varphi(u; t)$ , then the Fourier transform of this equation, that is, the Fokker–Planck equation for the stationary density  $f_s(x)$  of  $X(t)$  has the form

$$-\frac{d}{dx}((\rho x - x^2)f_s(x)) + \lambda \int \frac{1}{1 + \gamma y} f_s\left(\frac{x}{1 + \gamma y}\right) dF(y) - \lambda f_s(x) = 0, \quad (7.106)$$

where  $F$  denotes the distribution of  $Y_1$ .

*Proof* Calculations similar to those used for Theorem 7.4 give (7.106). For example, the first term on the right side of the stationary version of (7.104) multiplied by  $\exp(-iux)/(2\pi)$  and integrated over the real line gives

$$\begin{aligned} \frac{1}{2\pi} \int \rho u e^{-iux} \varphi'_s(u) du &= \frac{\rho}{2\pi} \left[ u e^{-iux} \varphi_s(u) \Big|_{-\infty}^{\infty} - \int \frac{d}{du} (u e^{-iux}) \varphi_s(u) du \right] \\ &= -\frac{\rho}{2\pi} \left[ \int e^{-iux} \varphi_s(u) du - ix \int e^{-iux} u \varphi_s(u) du \right] = -\rho f_s(x) - \rho x f'_s(x) \end{aligned}$$

using integration by parts and postulated boundary conditions for  $\varphi_s(u)$ . The integral term on the right side of (7.104) yields

$$\begin{aligned} \frac{1}{2\pi} \int e^{-iux} \left[ \lambda \int \varphi_s((u(1+\gamma y)) dF(y) \right] du \\ = \lambda \int \left[ \frac{1}{2\pi} \int e^{-ivx/(1+\gamma y)} \varphi_s(v) \frac{dv}{1+\gamma y} \right] dF(y) = \lambda \int \frac{1}{1+\gamma y} f_s\left(\frac{x}{1+\gamma y}\right) dF(y) \end{aligned}$$

by the change of variables  $v = u(1 + \gamma y)$  and Fubini's theorem.  $\blacktriangle$

### 7.5.3 Solution of Uncertain Dynamic Systems by SROMs

Methods for solving stochastic differential equations with random coefficients have been discussed in Sects. 7.3 and 7.4. In this section, we extend our considerations in Sect. 7.4.6 on the solution of this class of equations by stochastic reduced order models (SROMs). Numerical examples are presented to illustrate the construction of SROM-based solutions and assess their accuracy. Bounds are first established on the discrepancy between solutions of deterministic equations that have the same functional form but different coefficients. The bounds are then extended to characterize the accuracy of SROM-based solutions.

The solution of SDEs by extended stochastic reduced order models (ESROMs) and methods for solving SDEs by using these models are discussed in Sect. A.3. ESROM-based solutions for stochastic algebraic and partial differential equations are presented in the following two chapters.

#### 7.5.3.1 Bounds on Discrepancy Between Solutions of Deterministic Systems

Let  $X(t)$  and  $\tilde{X}(t)$  be  $\mathbb{R}^d$ -valued functions defined by

$$\begin{aligned} \dot{X}(t) &= f(t, X(t)), \quad t \geq 0, \\ \dot{\tilde{X}}(t) &= \tilde{f}(t, \tilde{X}(t)), \quad t \geq 0, \end{aligned} \tag{7.107}$$

where  $f, \tilde{f} : [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  are such that these equations have unique solutions. We establish bounds on the discrepancy between  $X(t)$  and  $\tilde{X}(t)$ .

Some notations are introduced prior to constructing these bounds. The logarithmic norm of an  $(d, d)$ -matrix  $Q$  with real-valued entries is

$$\mu(Q) = \lim_{\rho \downarrow 0} \frac{\|I + \rho Q\| - 1}{\rho}, \quad (7.108)$$

where  $I$  is the identity matrix and  $\|\cdot\|$  is a subordinate matrix norm. If  $\|\cdot\|$  is the matrix norm subordinated to the Euclidean norm, then  $\mu(Q)$  is the largest eigenvalue of matrix  $(Q + Q')/2$  ([58], Sect. I.10). Let  $\mu(\partial f(t, Z(t))/\partial X)$  be the logarithmic norm of the  $(d, d)$ -matrix  $\partial f(t, Z(t))/\partial X = \{\partial f_u(t, Z(t))/\partial z_v, u, v = 1, \dots, d\}$ , where  $Z(t)$  denotes a vector with coordinates in  $R(t) = \times_{i=1}^d [X_i(t) \wedge \tilde{X}_i(t), X_i(t) \vee \tilde{X}_i(t)] \subset \mathbb{R}^d$ , that exists by the mean value theorem ([59], Sect. 20). Let  $\ell(t)$  be a real-valued function of time such that  $\mu(\partial f(t, Z(t))/\partial X) \leq \ell(t)$ , and let

$$\delta(t) = \|\tilde{f}(t, \tilde{X}(t)) - f(t, \tilde{X}(t))\| \quad (7.109)$$

be a measure of the difference between functions  $\tilde{f}$  and  $f$  on the solution  $\tilde{X}(t)$ .

**Theorem 7.27** *The discrepancy between the solutions of the differential equations in (7.107) can be bounded by*

$$\|\tilde{X}(t) - X(t)\| \leq e^{L(t)} \left( \|\tilde{X}(0) - X(0)\| + \int_0^t e^{-L(s)} \delta(s) ds \right), \quad (7.110)$$

where  $L(t) = \int_0^t \ell(s) ds$  and  $\|\cdot\|$  denotes the Euclidean norm.

*Proof* If  $L(t) < 0$  at all times, then the right side of (7.110) can be used as a measure of the discrepancy between  $\tilde{X}(t)$  and  $X(t)$  provided  $\delta(t)$  is bounded. If  $\tilde{X}(t)$  and  $X(t)$  coincide at the initial time, then (7.110) becomes

$$\|\tilde{X}(t) - X(t)\| \leq e^{L(t)} \int_0^t e^{-L(s)} \delta(s) ds. \quad (7.111)$$

The measure  $m(t) = \|\tilde{X}(t) - X(t)\|$  of the discrepancy between  $\tilde{X}(t)$  and  $X(t)$  calculated at a later time  $t + \Delta t$ ,  $\Delta t > 0$ , satisfies the following inequality

$$\begin{aligned} m(t + \Delta t) &= \|\tilde{X}(t + \Delta t) - X(t + \Delta t)\| \\ &\leq \|\tilde{X}(t) + \dot{\tilde{X}}(t)\Delta t - X(t) - \dot{X}(t)\Delta t\| + O(\Delta t)^2 \\ &\leq \|\tilde{X}(t) - X(t) + \Delta t(\tilde{f}(t, \tilde{X}(t)) - f(t, X(t)))\| + O(\Delta t)^2 \\ &\leq \|\tilde{X}(t) - X(t) + \Delta t(f(t, \tilde{X}(t)) - f(t, X(t)))\| \\ &\quad + \Delta t \|\tilde{f}(t, \tilde{X}(t)) - f(t, \tilde{X}(t))\| + O(\Delta t)^2 \\ &= \left\| I + \frac{\partial f(t, Z(t))}{\partial X} \right\| m(t) + \Delta t \delta(t) + O(\Delta t)^2 \\ &\leq \max_{Z(t) \in R(t)} \left\| I + \frac{\partial f(t, Z(t))}{\partial X} \right\| m(t) + \Delta t \delta(t) + O(\Delta t)^2 \end{aligned}$$



for a small time step  $\Delta t > 0$ . An alternative form of this inequality is

$$\frac{m(t + \Delta t) - m(t)}{\Delta t} \leq \frac{1}{\Delta t} \left( \max_{Z(t) \in R(t)} \left\| I + \frac{\partial f(t, Z(t))}{\partial X} \right\| - 1 \right) m(t) + \delta(t) + O(\Delta t).$$

This inequality and Theorem 10.6 in [58] yield (7.110).  $\blacktriangle$

**Theorem 7.28** *If the differential equations (7.107) are linear, that is, they have the form*

$$\begin{aligned} \dot{X}(t) &= A(t)X(t) + Y(t), \quad t \geq 0, \\ \dot{\tilde{X}}(t) &= \tilde{A}(t)\tilde{X}(t) + \tilde{Y}(t), \quad t \geq 0, \end{aligned} \quad (7.112)$$

with initial states  $X(0) = X_0$  and  $\tilde{X}(0) = \tilde{X}_0$ , then

$$\| \tilde{X}(t) - X(t) \| \leq e^{L(t)} \left( \| \tilde{X}_0 - X_0 \| + \int_0^t e^{-L(s)} \delta(s) ds \right), \quad (7.113)$$

where

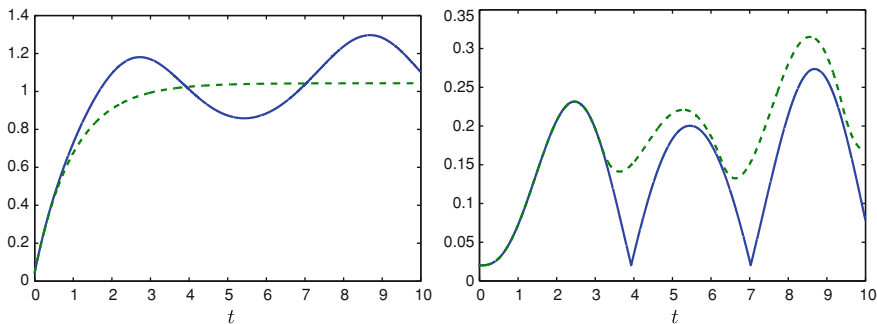
$$\begin{aligned} L(t) &= \int_0^t \lambda_{\max}(s) ds; \\ \lambda_{\max}(t) &= \text{the largest eigenvalue of } (A(t) + A(t)')/2; \quad \text{and} \\ \delta(t) &= \| (A(t) - \tilde{A}(t))\tilde{X}(t) + (Y(t) - \tilde{Y}(t)) \| . \end{aligned} \quad (7.114)$$

*Proof* Theorem 7.27 and the definition of the logarithmic norm yield the inequality in (7.113). If  $A$  is time invariant, then

$$\| \tilde{X}(t) - X(t) \| \leq e^{\lambda_{\max} t} \left( \| \tilde{X}_0 - X_0 \| + \int_0^t e^{-\lambda_{\max} s} \delta(s) ds \right), \quad (7.115)$$

where  $\lambda_{\max}$  is the largest eigenvalue of  $(A + A')/2$ . If  $\lambda_{\max} < 0$  and there exists a constant  $M^* > 0$  such that  $\delta(t) \leq M^*$  at all times, then  $\| \tilde{X}(t) - X(t) \| \leq \| \tilde{X}_0 - X_0 \| - M^*/\lambda_{\max}$  at all times. If  $\lambda_{\max} > 0$ , the resulting bound is less useful for our objective since it increases exponentially in time resembling the behavior of bounds derived from Gronwall's inequality ([60], Sect. 10.2).  $\blacktriangle$

*Example 7.48* Let  $X(t)$  and  $\tilde{X}(t)$  be real-valued functions defined by (7.112) with  $d = 1$ ,  $A(t) = -1 + 0.3 \sin(t)$ ,  $Y(t) = 1$ ,  $\tilde{A} = -1$ ,  $\tilde{Y}(t) = 1$ , and  $X(0) = \tilde{X}(0) = 0$ . We have  $\lambda_{\max}(t) = A(t)$ , so that  $\ell(t) = A(t)$  and  $L(t) = \int_0^t A(s) ds = -t + 0.3(1 - \cos(t))$ . Since  $\tilde{X}(t) = 1 - \exp(-t)$ , we have  $\delta(t) = |(A(t) - \tilde{A})\tilde{X}(t)| = |0.3 \sin(t)(1 - \exp(-t))|$ . The solid and dotted lines in Fig. 7.35 (left panel) are the solutions  $X(t)$  and  $\tilde{X}(t)$ . The solid and dotted lines in Fig. 7.35 (right panel) are the actual discrepancy  $|\tilde{X}(t) - X(t)|$  and the bound on  $|\tilde{X}(t) - X(t)|$  in (7.110). The bound is remarkably tight.  $\diamond$



**Fig. 7.35** Solutions  $X(t)$  and  $\tilde{X}(t)$  in solid and dotted lines (left panel), and  $|\tilde{X}(t) - X(t)|$  and a bound on  $|\tilde{X}(t) - X(t)|$  in solid and dotted lines (right panel)

### 7.5.3.2 SRM-Based Solutions

Suppose matrix  $A(t)$  and/or input  $Y(t)$  in (7.112) are random and that  $A(t) = A$  is time invariant. Let  $\mathcal{Z}$  be a random element collecting the random entries in the defining equation of  $X(t)$ , and assume that  $\mathcal{Z}$  can be characterized by  $n$  independent samples  $(z_1, \dots, z_n)$ . Let  $(\tilde{z}_1, \dots, \tilde{z}_m)$  and  $(p_1, \dots, p_m)$ ,  $m \ll n$ , be the defining parameters for a SRM  $\tilde{\mathcal{Z}}$  of  $\mathcal{Z}$ . Denote by  $\{X_i, A_i, Y_i\}$  and  $\{\tilde{X}_k, \tilde{A}_k, \tilde{Y}_k\}$  samples of  $(X, A, Y)$  corresponding to  $\{z_i\}$  and  $\{\tilde{z}_k\}$ . The discrepancy between  $X_i$  and  $\tilde{X}_k$  can be bounded by (Theorem 7.28)

$$\|\tilde{X}_k(t) - X_i(t)\| \leq e^{\lambda_{i,\max} t} \int_0^t e^{-\lambda_{i,\max} s} \delta_{k,i}(s) ds, \quad (7.116)$$

where  $\delta_{k,i}(t) = \|(A_i - \tilde{A}_k)\tilde{X}_k(t) + (Y_i(t) - \tilde{Y}_k(t))\|$  and  $\lambda_{i,\max}$  is the largest eigenvalue of  $(A_i + A_i')/2$ .

The inequalities in (7.116) can be used to construct bounds on moments of  $\|\tilde{X}(t) - X(t)\|$ . Let  $h$  be a measurable function defined by a partition  $\{\mathcal{C}_k\}$ ,  $k = 1, \dots, m$ , of  $(z_1, \dots, z_n)$  such that  $h(z_i) = \tilde{z}_k$  for  $z_i \in \mathcal{C}_k$  and  $p_k \simeq n_k/n$ , where  $n_k$  denotes the cardinality of  $\mathcal{C}_k$  (Sect. A.3). Accordingly, the moment of order  $q$  of the discrepancy  $\|\tilde{X}(t) - X(t)\|$  can be bounded by

$$\begin{aligned} E[\|\tilde{X}(t) - X(t)\|^q] &\leq \frac{1}{n} \sum_{k=1}^m \sum_{z_i \in \mathcal{C}_k} \left( e^{\lambda_{i,\max} t} \int_0^t e^{-\lambda_{i,\max} s} \delta_{k,i}(s) ds \right)^q \\ &\simeq \sum_{k=1}^m p_k \left[ \frac{1}{n_k} \sum_{z_i \in \mathcal{C}_k} \left( e^{\lambda_{i,\max} t} \int_0^t e^{-\lambda_{i,\max} s} \delta_{k,i}(s) ds \right)^q \right]. \end{aligned} \quad (7.117)$$

Similar bounds can be developed for other moments of  $\|\tilde{X}(t) - X(t)\|$ .

Numerical illustrations in the remainder of this section are based on developments in [61], and include dynamic systems with random coefficients and/or input.

*Example 7.49* Let  $X(t)$ ,  $t \geq 0$ , be defined by  $\dot{X}(t) = -\lambda X(t) + Y(t)^2$ , where  $\dot{Y}(t) = -\eta Y(t) + (2\eta)^{1/2} W(t)$ ,  $\lambda, \eta > 0$ , and  $W(t)$  is a Gaussian white noise viewed as the formal derivative of a Brownian motion  $B(t)$ . Then  $(X, Y)$  is a diffusion process satisfying the stochastic differential equation

$$d \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \begin{bmatrix} -\lambda X(t) + Y(t)^2 \\ -\eta Y(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ (2\eta)^{1/2} \end{bmatrix} dB(t). \quad (7.118)$$

It is assumed that  $Y(0)$  is independent of  $B$  and that  $Y(0) \sim N(0, 1)$ , so that  $Y(t)$  has a stationary start.

The linear random vibration theory can be applied to calculate second moment properties of  $X(t)$  defined by  $\dot{X}(t) = -\lambda X(t) + Y(t)^2$  since the mean and covariance functions of  $Z(t) = Y(t)^2$  can be calculated. For example, the stationary first two moments of  $X(t)$  are  $\mu(1, 0) = 1/\lambda$  and  $\mu(2, 0) = (3\lambda + 2\eta)/[\lambda^2(\lambda + 2\eta)]$ , so that its stationary variance is  $\sigma_{s,x}^2 = 2/[\lambda(\lambda + 2\eta)]$  (Examples 7.3 and 7.4).

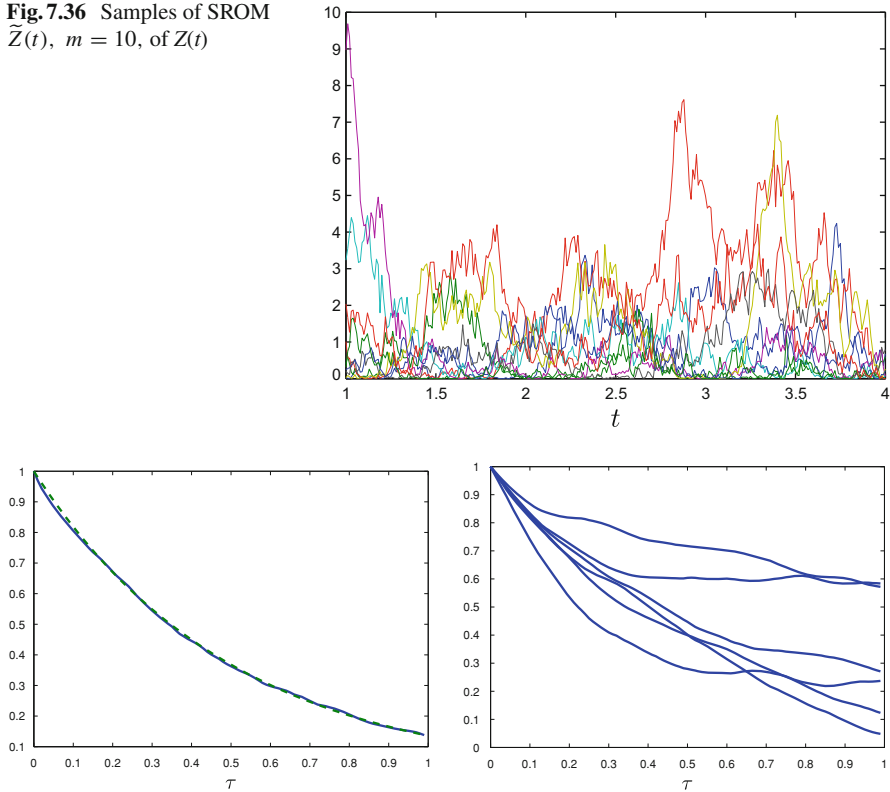
SROMs can be used to calculate approximately properties of  $X(t)$ . The following illustration is for  $\lambda = \eta = 1$  and a SROM  $\tilde{Z}(t)$  with  $m = 10$  samples consisting of samples of  $Z(t) = Y(t)^2$  generated in  $[0, 4]$ . Figure 7.36 shows the samples of  $\tilde{Z}(t)$  in the time interval  $[1, 4]$  in which  $Z(t)$  is assumed to be stationarity. The probabilities of these samples are  $p_1 = 0.0043$ ,  $p_2 = 0.0110$ ,  $p_3 = 0.1821$ ,  $p_4 = 0.2222$ ,  $p_5 = 0.0391$ ,  $p_6 = 0.0091$ ,  $p_7 = 0.3076$ ,  $p_8 = 0.0892$ ,  $p_9 = 0.1353$ , and  $p_{10} = 0.0001$ . The absolute value of the errors for the first six moments of  $\tilde{Z}(t)$  is under 2%. The solid and dotted lines in Fig. 7.37 (left panel) are the scaled covariance function of  $\tilde{Z}(t)$  and the scaled target covariance function  $c_z(\tau)/c_z(0)$ , where  $c_z(\tau) = 2\exp(-2\eta|\tau|)$  is the covariance function of  $Z(t) = Y(t)^2$ . Figure 7.37 (right panel) shows estimates of  $c_z(\tau)/c_z(0)$  obtained from sets of ten independent samples of  $Z(t)$ . In contrast to the SROM-based approximation for  $c_z(\tau)/c_z(0)$  with  $m = 10$  which is accurate, Monte Carlo estimates of  $c_z(\tau)/c_z(0)$  are unstable and can have large errors.

The absolute value of the errors for the first six moments of  $\tilde{X}(t)$  are 2.80, 2.96, 9.04, 16.02, 24.92, and 35.57 in percentages. These errors are calculated with respect to the exact solution of the moment equations for  $X(t)$ . While these errors are larger than those for  $\tilde{Z}(t)$ , the performance of  $\tilde{X}(t)$  is satisfactory given that this process has only  $m = 10$  samples. The solid and dotted lines in Fig. 7.38 (left panel) are the scaled covariance function of  $\tilde{X}(t)$  and the target covariance function  $c_x(\tau)/c_x(0)$ . Figure 7.38 (right panel) shows estimates of  $c_x(\tau)/c_x(0)$  obtained from sets of ten independent samples of  $Z(t)$ . The behavior of these estimates is similar to that of the estimates in Fig. 7.37 (right panel).

The accuracy of SROM-based solutions can be improved by increasing model size. For example, a SROM  $\tilde{X}(t)$  of  $X(t)$  corresponding to  $\tilde{Z}(t)$  with  $m = 14$  is superior to  $\tilde{X}(t)$  consider in previous figures. The covariance functions of  $\tilde{X}(t)$  with  $m = 14$  and  $m = 10$  are similar, but higher order moments of  $\tilde{X}(t)$  with  $m = 14$  are significantly more accurate than those of  $\tilde{X}(t)$  with  $m = 10$ . The absolute value of the errors of the first six moments of  $\tilde{X}(t)$  with  $m = 14$  are under 7.5%.

Temporal averages for properties of  $\tilde{X}(t)$  have been used in the previous evaluations for the following two reasons. First, simple processes, that is, processes having

**Fig. 7.36** Samples of SROM  $\tilde{Z}(t)$ ,  $m = 10$ , of  $Z(t)$



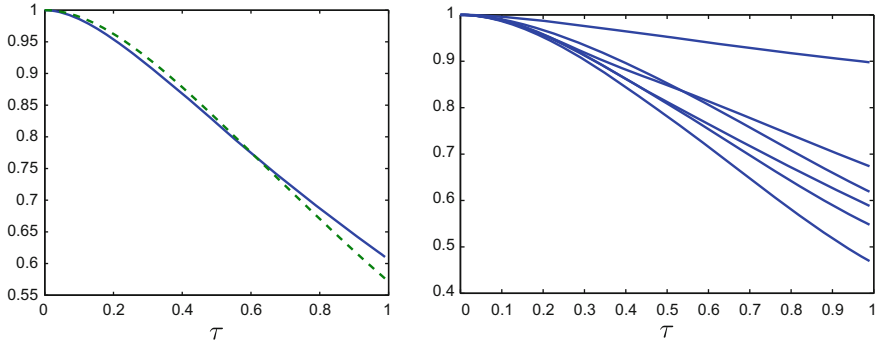
**Fig. 7.37** Scaled covariance functions of  $\tilde{Z}(t)$ ,  $m = 10$ , and  $Z(t)$  in solid and dotted lines (left panel) and Monte Carlo estimates of  $c_z(\tau)/c_z(0)$  for sets of ten samples of  $Z(t)$  (right panel)

a finite number of outcomes, cannot be stationary. This is consistent with the fact that estimates of stationary processes based on finite number of samples are not invariant to time shift. Second, temporal averages of statistics of simple processes are analogous to the assumption of random start. For example, the temporal average of the mean of  $\tilde{Z}(t)$  in a time interval  $[\tau_0, \tau_0 + \tau]$  is

$$\frac{1}{\tau} \int_{\tau_0}^{\tau_0+\tau} E[\tilde{Z}(t)] dt = \frac{1}{\tau} \int_{\tau_0}^{\tau_0+\tau} \sum_{k=1}^m p_k \tilde{z}_k(t) dt = \sum_{k=1}^m p_k \left( \frac{1}{\tau} \int_0^{\tau} \tilde{z}_k(\tau_0 + u) du \right)$$

where the latter integral can be interpreted as the expectation of  $\tilde{z}_k$  corresponding to a random uniformly distributed start in  $(0, \tau)$ .  $\diamond$

**Example 7.50** System responses  $X(t) = X_0 e^{-\lambda t} + \int_0^t e^{-\lambda(t-s)} Z(s) ds$  and  $\tilde{X}(t) = X_0 e^{-\lambda t} + \int_0^t e^{-\lambda(t-s)} \tilde{Z}(s) ds$  can be used directly to bound the discrepancy  $|X(t) - \tilde{X}(t)|$  between the processes in the previous example. The bound,



**Fig. 7.38** Scaled covariance functions of  $\tilde{X}(t)$ ,  $m = 10$ , and  $X(t)$  in solid and dotted lines (left panel) and Monte Carlo estimates of  $c_X(\tau)/c_X(0)$  for sets of ten samples of  $X(t)$  (right panel)

$$|\tilde{X}(t) - X(t)| \leq e^{-\lambda t} \int_0^t e^{-\lambda s} |\tilde{Z}(s) - Z(s)| ds,$$

derived from the expressions of  $X(t)$  and  $\tilde{X}(t)$  coincides with that in (7.115). Note also that

$$E[|\tilde{X}(t) - X(t)|] \leq e^{-\lambda t} \int_0^t e^{\lambda s} E[|\tilde{Z}(s) - Z(s)|] ds \leq \frac{M}{\lambda} (1 - e^{-\lambda \tau})$$

at any time  $t \in [0, \tau]$ , where  $M = \max_{0 \leq t \leq \tau} E[|\tilde{Z}(t) - Z(t)|]$ . If  $Z(t)$  is represented by  $n$  independent samples that are grouped in  $m$  clusters  $\mathcal{C}_k$ ,  $k = 1, \dots, m$ , centered on the samples  $\tilde{z}_k(t)$  of  $\tilde{Z}(t)$ , then

$$E[|\tilde{Z}(t) - Z(t)|] \simeq \frac{1}{n} \sum_{k=1}^m \sum_{z_i \in \mathcal{C}_k} |\tilde{z}_k(t) - z_i(t)|.$$

The graph of  $E[|\tilde{Z}(t) - Z(t)|]$  in Fig. 7.39 is based on  $n = 500$  independent samples of  $Z(t)$ . The corresponding constant  $M$  is approximately equal to 1.78 for  $\tau = 4$  and the SROM  $\tilde{Z}(t)$  of  $Z(t)$  with  $m = 10$  described in Figs. 7.36 and 7.37.  $\diamond$

**Example 7.51** Consider the stochastic differential equation

$$dX(t) = -\Lambda X(t) dt + dB(t), \quad t \geq 0, \quad (7.119)$$

where  $B(t)$  is a standard Brownian motion and  $\Lambda > 0$  is a random variable that is independent of  $B(t)$ . Let  $\tilde{\Lambda}$  be a SROM of  $\Lambda$  and denoted by  $\tilde{X}(t)$  the solution of (7.119) with  $\Lambda$  replaced by  $\tilde{\Lambda}$ . We develop bounds on the discrepancy between  $\tilde{X}(t)$  and  $X(t)$  using the fact that  $X(t) | \Lambda$  is a Gaussian process.

Numerical results are for  $\Lambda \sim U(1, 2)$ , a SROM  $\tilde{\Lambda}$  for  $\Lambda$  with  $m = 5$  samples, and  $X(0) = 1$ . The errors between the first six moments of  $\Lambda$  and  $\tilde{\Lambda}$  are under 2%. The solid and dotted lines in Fig. 7.40 (left panel) are the variance functions of  $\tilde{X}(t)$

and  $X(t)$ , respectively. Figure 7.40 (right panel) shows Monte Carlo estimates of the variance of  $X(t)$  based on sets of five independent samples of  $\Lambda$ . The Monte Carlo estimates show a significant sample to sample variation. The plots in Fig. 7.41 are similar to those in Fig. 7.40 but for central moments of order 4 of  $\tilde{X}(t)$  with  $m = 5$  and  $X(t)$ . These moments have the same behavior as the variances in Fig. 7.40. Monte Carlo estimates based on sets of five independent samples of  $\Lambda$  are unstable, and can be inaccurate.

The discrepancy between  $X(t)$  and  $\tilde{X}(t)$  can be measured by, for example,

$$E[|\tilde{X}(t) - X(t)|] \leq \frac{1}{n} \sum_{k=1}^m \sum_{\lambda_i \in \mathcal{C}_k} e^{-\lambda_i t} |\tilde{\lambda}_k - \lambda_i| \int_0^t e^{\lambda_i s} E[|\tilde{X}_k(s)|] ds, \quad (7.120)$$

where  $(\lambda_1, \dots, \lambda_n)$  are independent samples of  $\Lambda$  and  $(\tilde{\lambda}_k, p_k)$ ,  $k = 1, \dots, m$ , are the defining parameters for  $\tilde{\Lambda}$ . The sets  $\{\mathcal{C}_k, k = 1, \dots, m\}$  partition  $(\lambda_1, \dots, \lambda_n)$  such that  $p_k \simeq n_k/n$ , where  $n_k$  denotes the cardinality of  $\mathcal{C}_k$ . Since the Gaussian variable  $\tilde{X}_k(t)$  has mean 0 and variance  $\sigma_k(t)^2 = (1 - \exp(-2\tilde{\lambda}_k t))/(2\tilde{\lambda}_k)$ , we have  $E[|\tilde{X}_k(t)|] = \sqrt{2/\pi} \sigma_k(t)$  and

$$E|\tilde{X}(t) - X(t)| \leq \frac{1}{n\sqrt{2\pi}} \sum_{k=1}^m \sum_{\lambda_i \in \mathcal{C}_k} \frac{|\tilde{\lambda}_k - \lambda_i|}{\tilde{\lambda}_k} \left( \frac{1 - e^{-\lambda_i t}}{\lambda_i} - \frac{e^{-2\tilde{\lambda}_k t} - e^{-\lambda_i t}}{\lambda_i - 2\tilde{\lambda}_k} \right). \quad (7.121)$$

Similar considerations apply to the differential equations for the first two moments of  $\tilde{X}(t)$  and  $X(t)$ , that is, the means  $\dot{\mu}_i(t) = -\lambda_i \mu_i(t)$ ,  $\dot{\tilde{\mu}}_k(t) = -\tilde{\lambda}_k \tilde{\mu}_k(t)$ , the variances  $\dot{\gamma}_i(t) = -2\lambda_i \gamma_i(t) + 1$ ,  $\dot{\tilde{\gamma}}_k(t) = -2\tilde{\lambda}_k \tilde{\gamma}_k(t) + 1$ , and the covariance functions  $\partial c_i(s, t)/\partial t = -\lambda_i c_i(s, t)$ , and  $\partial \tilde{c}_k(s, t)/\partial t = -\tilde{\lambda}_k \tilde{c}_k(s, t)$ ,  $s < t$ , equations, where  $\mu_i$ ;  $\tilde{\mu}_k$ ,  $\gamma_i$ ;  $\tilde{\gamma}_k$ , and  $c_i$ ;  $\tilde{c}_k$  denote the mean, the variance, and the covariance functions corresponding to  $\Lambda$  set equal to  $\lambda_i$ ;  $\tilde{\lambda}_k$ . For example, the discrepancy between the mean functions  $\tilde{\mu}_k$  and  $\mu_i$  is such that

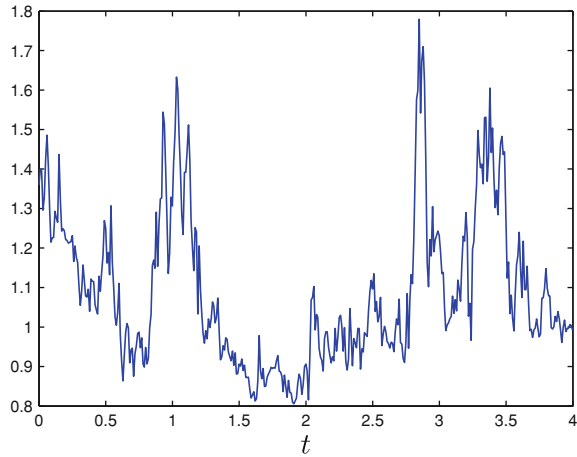
$$|\tilde{\mu}_k(t) - \mu_i(t)| \leq |\mu_0| \text{sign}(\tilde{\lambda}_k - \lambda_i) (e^{-\lambda_i t} - e^{-\tilde{\lambda}_k t})$$

by the bound in (7.115), where  $\mu_0$  denote the mean value of the initial state  $X(0)$ . The expectation of the discrepancy between the random functions  $\tilde{\mu}(t)$  and  $\mu(t)$  with outcomes  $\tilde{\mu}_k(t)$  of probabilities  $p_k$ ,  $k = 1, \dots, m$ , and  $\mu_i(t)$  of probability  $1/n$ ,  $i = 1, \dots, n$ , respectively, can be obtained from

$$E[|\tilde{\mu}(t) - \mu(t)|] \leq \frac{|\mu_0|}{n} \sum_{k=1}^m \sum_{\lambda_i \in \mathcal{C}_k} \text{sign}(\tilde{\lambda}_k - \lambda_i) (e^{-\lambda_i t} - e^{-\tilde{\lambda}_k t}).$$

Numerical results are for  $\tilde{\lambda}_k = 0.9 + 0.2k$ ,  $p_k = 1/m$ ,  $k = 1, \dots, m$ , and  $m = 5$ . Let  $I_k = [1 + 0.2(k-1), 1 + 0.2k)$  be a partition of the range  $[1, 2]$  of  $\Lambda$  and  $h: \mathbb{R} \rightarrow \mathbb{R}$  a measurable function such that  $h(I_k) = \tilde{\lambda}_k$  and  $P(\Lambda \in I_k) = p_k$ ,  $k = 1, \dots, m$ . If  $\Lambda \sim U(1, 2)$  is described by its distribution function rather than independent samples, we have

**Fig. 7.39** An estimate of  $E[|\tilde{Z}(t) - Z(t)|]$  for a SROM  $\tilde{Z}(t)$  with  $m = 10$



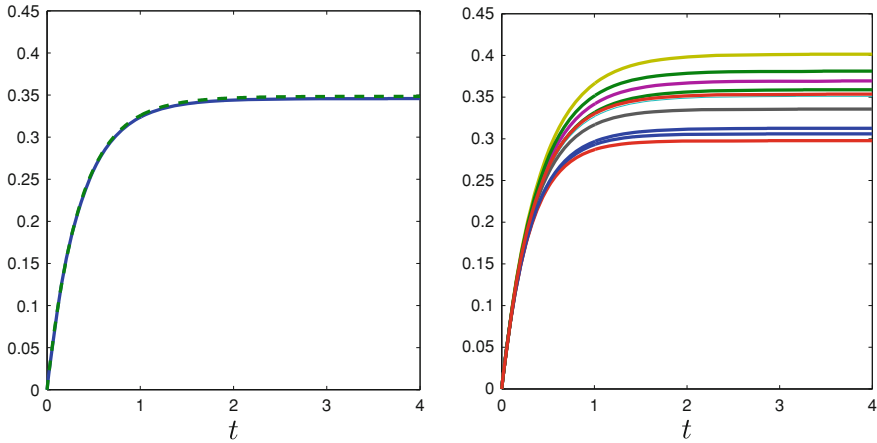
$$\begin{aligned}
 E[|\tilde{\mu}(t) - \mu(t)|] &\leq |\mu_0| \sum_{k=1}^m \int_{I_k} \text{sign}(\tilde{\lambda}_k - \lambda) (e^{-\lambda_i t} - e^{-\tilde{\lambda}_k t}) d\lambda \\
 &= \frac{|\mu_0|}{mt} \sum_{k=1}^m \left( e^{-(1+0.2(k-1))t} + e^{-(1+0.2k)t} - 2e^{-\tilde{\lambda}_k t} \right)
 \end{aligned}$$

Figure 7.42 shows this upper bound on  $E[|\tilde{\mu}(t) - \mu(t)|]$  in  $[0,4]$  for  $|\mu_0| = 1$ .  $\diamond$

*Example 7.52* Suppose the coefficient  $\lambda$  in Example 7.49 is a real-valued random variable  $\Lambda > 0$ , so that  $X(t)$  satisfies a stochastic equation with random coefficients driven by a non-Gaussian process  $Z(t) = Y(t)^2$ . Let  $\tilde{Z}(t)$  be a SROM for  $Z(t)$  with samples  $(\tilde{z}_1, \dots, \tilde{z}_m)$  of probabilities  $(p_1, \dots, p_m)$  and let  $\tilde{\Lambda}$  be a SROM for  $\Lambda$  with samples  $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{\tilde{m}})$  of probabilities  $(q_1, \dots, q_{\tilde{m}})$ , where  $m$  may or may not be equal to  $\tilde{m}$ . Numerical results are for  $X(t)$  in stationary regime,  $\eta = 1$ ,  $\Lambda \sim U(1, 2)$ ,  $m = 10$ , and  $\tilde{m} = 5$ , so that the SROM  $\tilde{X}(t)$  for  $X(t)$  has 50 samples of probabilities  $p_k q_l$ ,  $k = 1, \dots, m$ ,  $l = 1, \dots, \tilde{m}$ . As previously, temporal averages of properties of  $\tilde{X}(t)$  are compared with corresponding properties of  $X(t)$ .

The absolute value of the errors of the first six moments of  $\tilde{X}(t)$  relative to Monte Carlo estimates of these moments based on 1000 samples are 3.43, 1.08, 5.74, 12.10, 21.02, and 31.97 in percentages. The solid and dotted lines in Fig. 7.43 (left panel) are the scaled covariance functions of  $\tilde{X}(t)$  and  $X(t)$  for time lags in the range  $[0,1]$ . Fig. 7.43 (right panel) shows Monte Carlo estimates of the scaled covariance function  $c_X(\tau)/c_X(0)$  of  $X(t)$  based on 5 samples of  $\Lambda$  and 10 samples of  $Z(t)$ . The estimates exhibit notable sample to sample variation, and can have large errors. The accuracy of  $\tilde{X}(t)$  can be improved by increasing its size. For example, the absolute errors of the first six moments of  $\tilde{X}(t)$  corresponding to  $\tilde{Z}(t)$  and  $\tilde{\Lambda}$  with  $m = 14$  and  $\tilde{m} = 5$  do not exceed 8%.

As in the previous examples, properties of  $\tilde{X}(t)$  can be obtained by elementary calculations. For example, the moment of order  $r \geq 1$  at a time  $t$  and the correlation



**Fig. 7.40** Variance functions of  $\tilde{X}(t)$ ,  $m = 5$ , and  $X(t)$  in solid and dotted lines (left panel) and Monte Carlo estimates of the variance function of  $X(t)$  for sets of five samples of  $X(t)$  (right panel)

function of  $\tilde{X}(t)$  at times  $s$  and  $t$  are

$$\begin{aligned} E[\tilde{X}(t)^r] &= \sum_{l=1}^{\bar{m}} \left( \sum_{k=1}^m (\tilde{x}_{k,l}(t))^r p_k \right) q_l \\ E[\tilde{X}(s)\tilde{X}(t)] &= \sum_{l=1}^{\bar{m}} \left( \sum_{k=1}^m (\tilde{x}_{k,l}(s)\tilde{x}_{k,l}(t))^r p_k \right) q_l, \end{aligned} \quad (7.122)$$

where  $\tilde{X}_{k,l}(t)$  is the solution of (7.118) with  $\Lambda$  and  $Z(t)$  set equal to  $\tilde{\lambda}_l$  and  $\tilde{z}_k(t)$ , respectively.  $\diamond$

*Proof* The discrepancy between the solutions  $\tilde{X}_{k,l}(t)$  and  $X_{i,j}(t)$  defined by (7.118) with  $(\Lambda, Z(t))$  replaced with  $(\tilde{\lambda}_l, \tilde{z}_k(t))$  and  $(\lambda_j, z_i(t))$  is

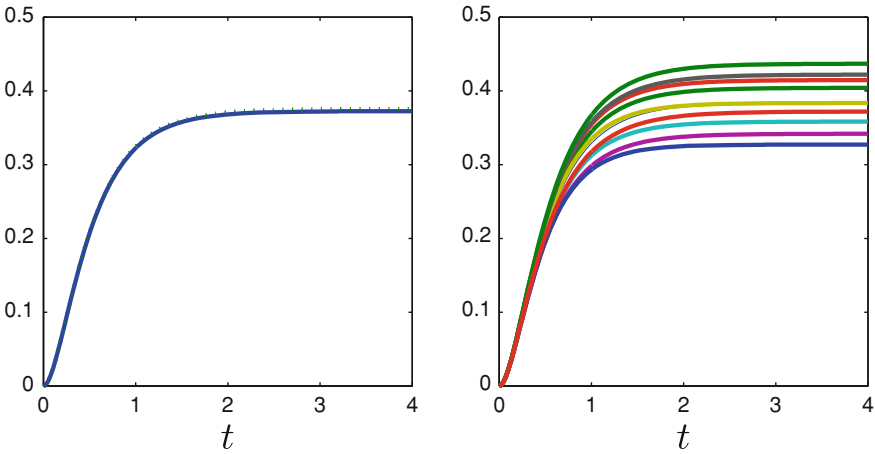
$$|\tilde{X}_{k,l}(t) - X_{i,j}(t)| \leq e^{-\lambda_j t} \int_0^t e^{\lambda_j s} \delta_{i,j,k,l}(s) ds,$$

where  $\delta_{i,j,k,l}(s) = |(\lambda_j - \tilde{\lambda}_l) \tilde{x}_{k,l}(s) + (z_i(s) - \tilde{z}_k(s))|$ . These inequalities can be used to bound moments of  $|\tilde{X}(t) - X(t)|$ . For example,

$$E[|\tilde{X}(t) - X(t)|] \leq \frac{1}{\bar{n}} \sum_{l=1}^{\bar{m}} \sum_{\lambda_j \in \mathcal{L}_l} \left( \frac{1}{n} \sum_{k=1}^m \sum_{z_i \in \mathcal{C}_k} e^{-\lambda_j t} \int_0^t e^{\lambda_j s} \delta_{i,j,k,l}(s) ds \right), \quad (7.123)$$

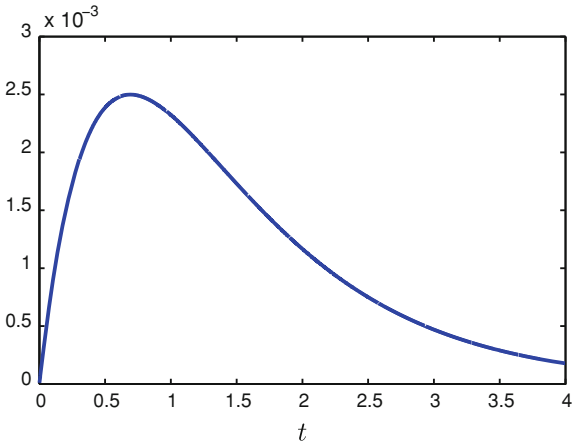
where  $\mathcal{C}_k$ ,  $k = 1, \dots, m$ , and  $\mathcal{L}_l$ ,  $l = 1, \dots, \bar{m}$ , denote partitions of the set of independent samples  $\{z_i(t)\}$  and  $\{\lambda_j\}$  used to represent  $Z(t)$  and  $\Lambda$ .  $\blacktriangle$





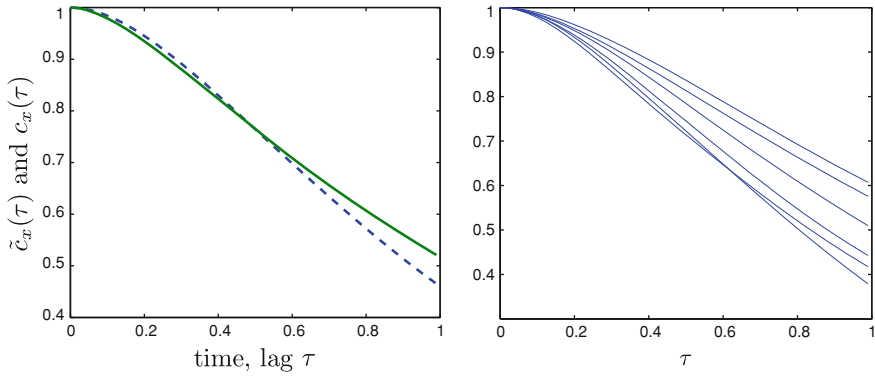
**Fig. 7.41** Central moments of order 4 of  $\tilde{X}(t)$ ,  $m = 5$ , and  $X(t)$  in solid and dotted lines (left panel) and Monte Carlo estimates of the central moment of order 4 of  $X(t)$  for sets of five samples of  $X(t)$  (right panel)

**Fig. 7.42** Bound on  $E[|\tilde{\mu}(t) - \mu(t)|]$  for  $m = 10$  and  $n=500$



**7.5.4 Degrading Systems**

Suppose a through rectilinear crack of length  $2a_0$  is detected in the wing of an aircraft. Our objective is to find the probability that the crack length does not exceed a critical value  $a_{cr} > 0$  during the time  $\tau$  between scheduled inspections. It is assumed that the action on the wing can be modeled by a stationary broad band Gaussian process, the crack does not affect the overall wing behavior, the wing is a linear system whose dynamics can be captured by its first mode of vibration, and the stress process  $X(t)$  normal to the crack controls its growth. Under these assumptions,  $X(t)$  is the solution of



**Fig. 7.43** Scaled covariance functions of  $\tilde{X}(t)$ ,  $m = 10$  and  $\bar{m} = 5$ , and  $X(t)$  in solid and dotted lines (left panel) and Monte Carlo estimates of  $c_x(\tau)/c_x(0)$  for sets of five independent samples of  $\Lambda$  and ten independent samples of  $X(t)$  (right panel)

$$\ddot{X}(t) + 2\zeta v_0 \dot{X}(t) + v_0^2 X(t) = \varepsilon W(t), \quad t \geq 0, \quad (7.124)$$

where  $W(t)$  is a Gaussian white noise with mean 0 and constant spectral density of magnitude  $s_0$ ,  $\varepsilon > 0$  is a small parameter,  $v_0 > 0$  denotes the frequency of the first mode, and  $\zeta = \varepsilon^2 \in (0, 1)$  is the damping ratio in this mode. For a sufficiently large time,  $X(t)$  is a stationary narrow band Gaussian process with mean 0, variance  $\sigma_{st}^2 = \pi s_0 / (2v_0^3)$ , and spectral density  $s(v) = s_0 / [(v^2 - v_0^2)^2 + (2\zeta v v_0)^2]$ , that admits the representation

$$X(t) = H(t) \cos(v_0 t + \Psi(t)), \quad (7.125)$$

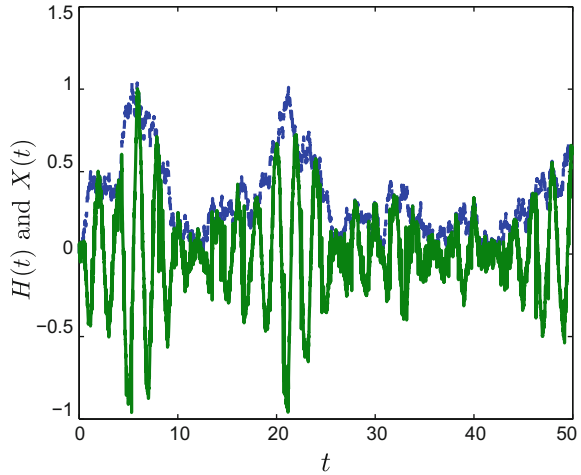
that is,  $X(t)$  is an harmonic with frequency  $v_0$  and random phase  $\Psi(t)$  that is modulated by a random amplitude  $H(t)$  ([62], Sect. 14.4, [20], Examples 5.5 and 5.12). The oscillations of the stress process  $X(t)$  are at a time scale much shorter than those of  $H(t)$  and  $\Psi(t)$ . It can be shown that the amplitude  $H(t)$  and the phase  $\psi(t)$  of  $X(t)$  are diffusion processes defined by the stochastic differential equations

$$\begin{aligned} dH(t) &= -\varepsilon^2 \left( v_0 H(t) - \frac{\pi s_0}{2v_0^2 H(t)} \right) dt + \varepsilon \frac{\sqrt{\pi s_0}}{v_0} dB_1(t) \\ d\Psi(t) &= \varepsilon \frac{\sqrt{\pi s_0}}{v_0 H(t)} dB_2(t) \end{aligned} \quad (7.126)$$

driven by the independent Brownian motions  $B_1(t)$  and  $B_2(t)$  [52]. The first equation takes the form

$$dR(t) = -\varepsilon^2 \left( R(t) - \frac{1}{2R(t)} \right) dt + \varepsilon \sqrt{v_0} dB_1(t), \quad (7.127)$$

**Fig. 7.44** Samples of  $X(t)$  and  $H(t)$  for  $\varepsilon = 0.3$ ,  $s_0 = 1$ ,  $v_0 = \pi$



by the change of variables  $R(t) = H(t)/(\sqrt{2}\sigma_{st})$ . Figure 7.44 shows a samples of  $H(t)$  generated from (7.126) and the sample of  $X(t)$  calculated from (7.125) using samples of  $H(t)$  and  $\Psi(t)$ .

The Paris–Erdogan model predicts that a crack of length  $a$  extends by an amount  $\alpha(\eta(A)\sqrt{\pi A}(2H))^\beta$  during a cycle of the far stress field with range  $2H$ , where  $\alpha$ ,  $\beta$  are material constants and  $\eta$  is a function of specimen geometry ([20], Sect. 7.5.2). Since the duration of stress cycles is  $2\pi/v_0$ , a crack with length  $A(t)$  increases at a rate

$$\begin{aligned}\dot{A}(t) &= \frac{v_0}{2\pi} \alpha \left( \eta(A(t)) \sqrt{\pi A(t)} 2H(t) \right)^\beta \\ &= \frac{v_0}{2\pi} \alpha \left( \eta(A(t)) \sqrt{\pi A(t)} 2\sqrt{2}\sigma_{st} R(t) \right)^\beta = g(A(t), R(t)),\end{aligned}\quad (7.128)$$

so that  $(A(t), R(t))$  is an  $\mathbb{R}^2$ -valued diffusion process satisfying the Itô stochastic differential equations

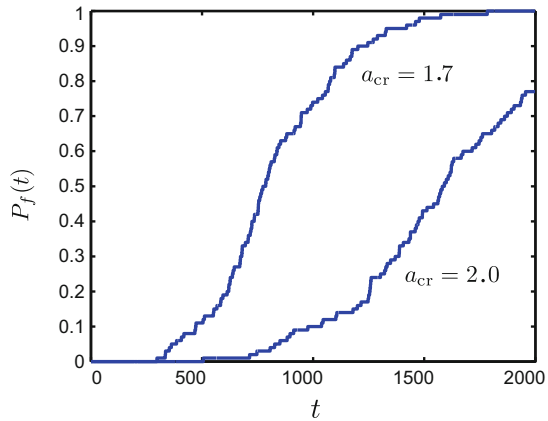
$$\begin{cases} dA(t) = g(A(t), R(t))dt \\ dR(t) = -\varepsilon^2 \left( R(t) - \frac{1}{2R(t)} \right) dt + \varepsilon \sqrt{v_0} dB_1(t). \end{cases}\quad (7.129)$$

**Example 7.53** Suppose that the initial crack length  $A(0) = a_0 < a_{cr}$  is deterministic and known and that our objective is to calculate system reliability  $P_s(t)$ , that is, the probability that the crack length does not exceed a critical value  $a_{cr} > 0$  during a time interval  $[0, t]$ ,  $t \geq 0$ .

An alternative form of (7.128) is  $d\chi(A(t)) = R(t)^\beta dt$ , where

$$d\chi(a) = \frac{c da}{\eta(a)^\beta a^{\beta/2}} \quad \text{and} \quad c = \frac{2\pi}{\alpha(\sqrt{2}\sigma_{st})^\beta}.$$

**Fig. 7.45** Failure probabilities  $P_f(t)$  for  $a_{cr} = 1.7$  and  $2.0$  in



This gives  $\chi(A(t)) - \chi(A(0)) = \int_0^t R(s)^\beta ds$  and  $A(t) = \chi^{-1}(\chi(A(0)) + \bar{R}(t))$ , so that the failure probability  $P_f(t) = 1 - P_s(t)$  can be calculated from

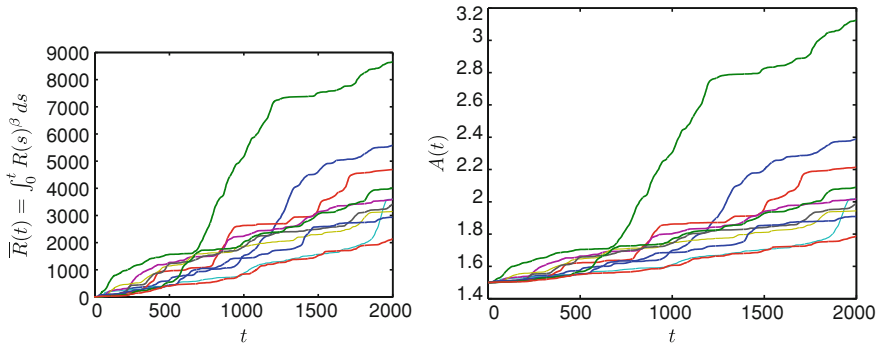
$$P_f(t) = P(A(t) > a_{cr}) = P(\bar{R}(t) > \chi(a_{cr}) - \chi(a_0)),$$

where  $\bar{R}(\tau) = \int_0^\tau R(s)^\beta ds$ . Additional information on stochastic fatigue crack growth can be found in [20] (Sect. 7.5.2) and [63] (Sect. 9.2).

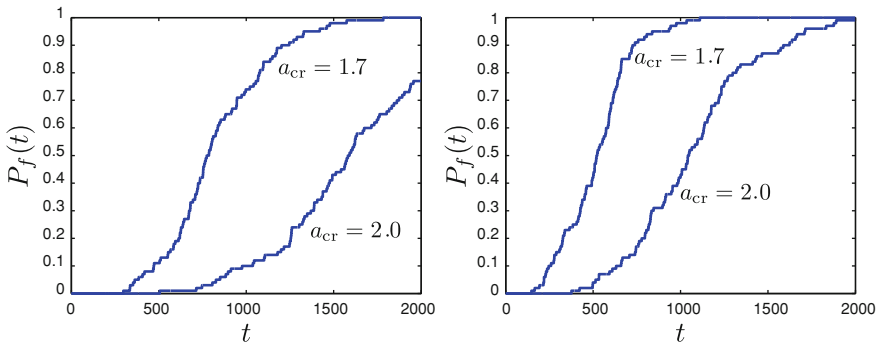
Figure 7.45 shows estimates of failure probabilities  $P_f(t) = 1 - P_s(t)$  for  $\alpha = 6.6 \times 10^{-7}$ ,  $\beta = 2.25$ ,  $s_0 = 100$ ,  $v_0 = \pi$  rad/s,  $a_0 = 1.5$  in, and two critical crack lengths,  $a_{cr} = 1.7$  and  $2.0$  in. The estimates of  $P_f(t)$  have been obtained from 100 independent samples of  $A(t)$ . The left and the right panels of Fig. 7.46 show ten samples of  $\bar{R}(\tau)$  and  $A(t)$ .  $\diamond$

The model in Example 7.53 assumes that system dynamics is not affected by damage, that is, crack length. A possible consequence of system degradation is that its modal frequency  $v_0$  in (7.124) becomes a function of crack length  $A(t)$ . As a result,  $X(t)$  would become a nonstationary, non-Gaussian process. A heuristic solution to this problem can be developed by using the observation that the temporal scales of  $A(t)$  and  $X(t)$  differ significantly. For example, we may assume that  $A(t)$  is constant over time intervals covering a large number of cycles of  $X(t)$  and that these intervals are much larger than the duration of transients of  $X(t)$  following a change in model frequency. Under these assumptions, the reliability analysis in Example 7.53 can be employed in the intervals of constant values of  $A(t)$ .

**Example 7.54** Suppose  $v_0$  in (7.124) decreases suddenly to  $v^* > 0$  at a time  $t^* \in (0, \tau)$  and that  $X(t)$  is stationary in  $(t^*, \tau)$ . The crack growth rates in the time intervals  $[0, t^*)$  and  $[t^*, \tau)$  are  $\dot{A}(t) = (1/c)\eta(A(t))^\beta A(t)^{\beta/2} R(t)^\beta$  and  $\dot{A}(t) = (1/c^*)\eta(A(t))^\beta A(t)^{\beta/2} R^*(t)^\beta$ , where  $c^* = 2\pi/[v^*\alpha(2\sqrt{2\pi}\sigma_{st}^*)^\beta]$ ,  $R^*(t)$  is given by (7.127) with  $v^*$  in place of  $v_0$ , and  $\sigma_{st}^* = \pi s_0/[2(v^*)^3]$ . The left panel in



**Fig. 7.46** Samples of  $\bar{R}(t)$  (left panel) and  $A(t)$  (right panel)



**Fig. 7.47** Estimates of  $P_f(t)$  for  $v_0$  (left panel) and  $v^* = 0.8v_0$  (right panel)

Fig. 7.47 shows the estimates of  $P_f(t)$  in Fig. 7.45. The plots in the right panel are estimates of  $P_f(t)$  for  $v^* = 0.8v_0$  obtained from 100 independent samples of  $A(t)$  and  $R(t)$ . The plots in the two panels of the figure show that modal frequency can affect significantly our estimate of  $P_f(\tau)$ .  $\diamond$

## 7.6 Exercises

**Exercise 7.1** Let  $X(t)$  be a real-valued process satisfying the differential equation  $\ddot{X}(t) + 2\zeta v_0 \dot{X}(t) + v_0^2 X(t) = W(t)$ ,  $t \geq 0$ , with initial conditions  $X(0) = 0$  and  $\dot{X}(0) = 0$ , where  $\zeta \in (0, 1)$ ,  $v_0 > 0$ , and  $W(t)$  denotes a real-valued white noise with mean 0 and one-sided spectral density  $g(v) = g_0 > 0$ ,  $v \geq 0$ . Find the expression of the covariance matrix  $\gamma_{X, \dot{X}}(t)$  of the bivariate vector  $(X(t), \dot{X}(t))$  as a function of time. Show that the entries (1,1), (2,2), and (1,2) of the stationary covariance matrix of  $(X(t), \dot{X}(t))$  are  $\pi g_0 / (4\zeta v_0^3)$ ,  $\pi g_0 / (4\zeta v_0)$ , and zero, respectively.

**Exercise 7.2** Complete the proof of Theorem 7.2 by showing that the correlation function of  $X(t)$  satisfies the third equation in (7.6).

*Hint* Multiply the equality  $X_p(t) - X_p(s) = \int_s^t dX_p(u)$ ,  $t > s$ , by  $X_q(s)$ , calculate the expectation of the resulting formula, and take its derivative with respect to time.

**Exercise 7.3** Extend the proof of Theorem 7.2 to the case in which  $B(t)$  has dependent coordinates. For example, set  $B(t) = \beta(t)B^*(t)$ , where  $\beta(t)$  is a  $(d', d^*)$  matrix with real-valued entries and  $B^*(t)$  denotes an  $\mathbb{R}^{d^*}$ -valued Brownian motion with independent coordinates.

**Exercise 7.4** Extend Theorem 7.2 to the case  $S(t) = C(t) = \sum_{k=1}^{N(t)} Y_k$ , where  $N(t)$  is a homogeneous Poisson process with intensity  $\lambda > 0$  and  $\{Y_k\}$  are iid  $\mathbb{R}^{d'}$ -valued random variables.

*Hint* Use Itô's formula in (5.16).

**Exercise 7.5** Calculate the second moment properties of  $Z(t) = Y(t)^k$ , where  $Y(t)$  is the process in Example 7.3.

**Exercise 7.6** Derive the differential equation for the covariance function of the state of a linear system driven by the colored noise given by (7.14).

**Exercise 7.7** Calculate and plot the correlation function of  $X(t)$  in Example 7.4 for  $k = 3$ .

**Exercise 7.8** Repeat the calculations in Exercise 7.7 with  $B(t)$  replaced by a compound Poisson process  $C(t)$  with the same second moment properties as  $B(t)$ .

**Exercise 7.9** Derive the partial differential equations for the characteristic and density functions for the diffusion processes in Example 7.11.

*Hint* The differential equations for the characteristic and distribution functions of  $X(t)$  driven by Gaussian white noise can be obtained directly from (7.17) and (7.18). For  $X(t)$  driven by Poisson white noise, apply Itô's formula. Use the independence between  $X(s-)$  and  $\Delta C(s) = C(s) - C(s-)$  and the fact that  $C(s + \Delta s) - C(s)$  for small  $\Delta s > 0$  is 0 or  $Y_k$  for some  $k$  with probabilities  $\lambda \Delta s$  or  $1 - \lambda \Delta s$ .

**Exercise 7.10** Calculate the mean and covariance of  $X_n$  conditional on a sample  $(T_0 = 0, T_1(\omega), \dots)$  of the jump times  $(T_0, T_1, \dots)$  of semi-Markov sequence in Example 7.17.

**Exercise 7.11** Derive the expression of  $X_{n+1}$  in (7.39).

**Exercise 7.12** Prove the last equality in (7.44) and find the second moment properties of  $X_{n,0} + \varepsilon X_{n,1}$ .

**Exercise 7.13** Show that  $\mathscr{W}$  given by (7.46) is a linear space, (7.47) defines an inner product on  $\mathscr{W}$ , and  $\mathscr{W}$  with the norm induced by this inner product is a Hilbert space.

**Exercise 7.14** Find the moment and the Fokker–Planck equations for the process in (7.76), that are given in Example 7.29.

**Exercise 7.15** Develop bounds as in Sect. 7.5.3.2 on the error of SROM-based solutions for simple linear oscillators and multi-degree of freedom linear systems with random stiffness and damping that are subjected to Gaussian white noise.

**Exercise 7.16** Calculate the correlation function of  $\tilde{X}(t)$  in Example 7.36 and establish conditions under which it converges to the correlation function of  $X(t)$  as  $m \rightarrow \infty$ .

## References

1. Bellman R, Kashef BG, Casti J (1972) Differential quadrature: A technique for the rapid solution of nonlinear partial differential equations. *J Comput Phys* 10:40–52
2. Johnson C (1994) Numerical solution of partial differential equations by the finite element method. Cambridge University Press, New York
3. Roth Ch (2002) Difference methods for stochastic partial differential equations. *ZAMM-Z Angew Math Mech* 82:821–830
4. Grigoriu M (2002) Stochastic calculus. Applications in science and engineering. Birkhäuser, Boston
5. Csörgő S, Hatvani L (2010) Stability properties of solutions of linear second order differential equations with random coefficients. *J Differ Equ* 248:21–49
6. Xie WC (2006) Dynamic stability of structures. Cambridge University Press, New York
7. Dzhalladova IA (1998) Investigation of a system of linear differential equations with random coefficients. *Ukrainian Math J* 50(8):1299–1307
8. Lapshin AL (1999) Equations for second moments of solutions of a system of linear differential equations with random semi-Markov coefficients and random input. *Ukrainian Math J* 51(6):864–873
9. Valeev KG, Dzhalladova IA (2002) Derivation of moment equations for solutions of a system difference equations dependent on a semi-Markov process. *Ukrainian Math J* 54(11):1906–1911
10. Valeev KG, Dzhalladova IA (2003) Derivation of moment equations for solutions of a system nonlinear difference equations dependent on a semi-Markov process. *Ukrainian Math J* 55(6):1043–1041
11. Valeev KG, Dzhalladova IA (2004) Second-order moment equations for a system of differential equations with random right-hand side. *Ukrainian Math J* 56(5):830–834
12. Valeev KG, Khalil GS (1997) Investigation of the solutions of linear systems of difference equations with random coefficients. *Period Math Hung* 35(1-2):139–144
13. Diaz J (1990) Bayesian forecasting for AR(1) models with normal coefficients. *Commun Stat Theory Methods* 19:229–2246
14. Nicholls DF (1986) The Box-Jenkins approach to random coefficients autoregressive modelling. *J Appl Probab* 23:231–240
15. Priestley MB (1988) Non-linear and non-stationary time series analysis. Academic Press, New York
16. Jacobs PA, Lewis PAW (1978) Discrete time series generated by mixtures I: Correlation and runs properties. *J Roy Stat Soc* 40:94–105
17. Grigoriu M (1995) Applied non-Gaussian processes. Examples, theory, simulation, linear random vibration and MATLAB solutions. Prentice Hall, Englewoods Cliffs
18. Grigoriu M (1990) Reliability analysis of dynamic Daniels systems with local load sharing rule. *J Eng Mech* 116:2625–2642
19. Rychlik I, Grigoriu M (1992) Reliability of Daniels systems with equal load sharing rule subject to stationary Gaussian dynamic loads. *Probab Eng Mech* 7:113–121

20. Soong TT, Grigoriu M (1993) Random vibration of mechanical and structural systems. Prentice-Hall, Englewood Cliffs
21. Tarp-Johansen NJ, Ditlevsen O (2001) Time between plastic displacements of elastic-plastic oscillators subject to Gaussian white noise. *Probab Eng Mech* 16:373–380
22. Mantegna RN, Stanley HE (2000) An introduction to econophysics correlation and complexity in finance. Cambridge University Press, Cambridge
23. Melsa JL, Sage AP (1973) An introduction to probability and stochastic processes. Prentice-Hall, Englewood Cliffs
24. Brockett RW (1970) Finite dimensional linear systems. Wiley, New York
25. Apostol TM (1974) Mathematical analysis. Addison-Wesley Publishing Company, Reading
26. Grigoriu M (2008) A critical evaluation of closure methods via two simple dynamic systems. *J Sound Vib* 317(1–2):190–198. doi:[10.1016/j.jsv.2008.02.049](https://doi.org/10.1016/j.jsv.2008.02.049)
27. Hasofer AM, Grigoriu M (1995) A new perspective on the moment closure method. *J Appl Mech* 62(2):527–532
28. Fung YC (1965) Foundations of solid mechanics. Prentice-Hall, Englewood Cliffs
29. Resnick SI (1992) Adventures in stochastic processes. Birkhäuser, Boston
30. Gardiner CW (1985) Handbook of stochastic methods for physics chemistry and the natural sciences, 2nd edn. Springer, New York
31. Roberts JB, Spanos PD (1990) Random vibration and statistical linearization. Wiley, New York
32. Brockwell PJ, Davis RA (1987) Time series: Theory and methods. Springer, New York
33. Leipus R, Surgailis D (2003) Random coefficients autoregression, regime switching and long memory. *Adv Appl Probab* 35:737–754
34. Abraham B, ChE Minder (1982) A time series model with random coefficients. *Commun Stat Theory Methods* 11(12):1381–1391
35. Grigoriu M, Veneziano D, Cornell CA (1979) Probabilistic modelling as decision making. *J Eng Mech Div ASCE* 105(EM4):585–596
36. Field RV (2004) Methods for model selection in applied science and engineering. PhD thesis, Cornell University, Ithaca, NY, May (Grigoriu M, thesis supervisor)
37. Field RV, Grigoriu M (2007) Model selection in applied science and engineering. A decision-theoretic approach. *J Eng Mech* 133(7):780–791
38. Zellner A (1971) An introduction to Bayesian inference in econometrics. Wiley, New York
39. Raiffa H, Schlaifer R (1961) Applied statistical decision theory. The MIT Press, Cambridge
40. Babuška I (1971) Error bounds for finite element method. *Numerische Mathematik*,
41. Kloeden PE, Platen E (1992) Numerical solutions of stochastic differential equations. Springer, New York
42. Cramer H, Leadbetter MR (1967) Stationary and related stochastic processes. Wiley, New York
43. Billingsley P, Probability and measure, 3rd edn. Wiley, New York
44. Lukacs E (1960) Characteristic functions. Number 5 in Griffin's statistical monographs & courses. Charles Griffin & Company Limited, London
45. Babuška IM, Nobile F, Tempone R (2007) A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J Numer Anal* 45(3):1005–1034
46. Foo J, Wan X, Karniadakis E (2008) The multi-element probabilistic collocation method (ME-PCM): Error analysis and applications. *J Comput Phys* 227:9572–9595
47. Nobile F, Tempone R, Webster CG (2008) An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J Numer Anal* 46(5):2411–1442
48. Hildebrand FB (1965) Methods of applied mathematics. Prentice Hall, Englewood Cliffs
49. Kanwal RP (1971) Linear integral equations. Theory and technique. Academic Press, New York
50. Tricomi FG (1957) Integral equations. Dover Publications, New York
51. Ariaratnam ST, Abdelrahman NM (2001) Almost sure stochastic stability of viscoelastic plates in supersonic flow. *AIAA J* 39(3):465–472



52. Ariaratnam ST, Pi HN (1973) On the first-passage time for envelope crossing for a linear oscillator. *Int J Control* 18(1):89–96
53. Ariaratnam ST, Xie WC (1994) Almost-sure stochastic stability of coupled non-linear oscillators. *Int J Non-Linear Mech* 29(2):197–204
54. Sidney Resnick (1992) *Adventures in stochastic processes*. Birkhäuser, Boston
55. Billingsley P (1968) *Convergence of probability measures*. Wiley, New York
56. Samorodnitsky G, Taqqu MS (1994) *Stable non-Gaussian random processes. stochastic models with infinite variance*. Birkhäuser, New York
57. Grigoriu M (2009) Numerical solution of stochastic differential equations with Poisson and Lévy white noise. *Phys Rev E* 80. doi:[10.1103/PhysRevE.80.026704](https://doi.org/10.1103/PhysRevE.80.026704)
58. Hairer E, Nørsett SP, Wanner G (1993) *Solving ordinary differential equations I. Nonstiff Problems*, Second Revised Edition. Springer, New York
59. Brabenec RL (1990) *Introduction to real analysis*. PWS-KENT Publishing Company, Boston
60. Hui-Hsiung Kuo, *Introduction to stochastic integration*. Springer, New York
61. Grigoriu M (2010) Linear random vibration by stochastic reduced order models. *Int. J Num Methods Eng* 82:1537–1559. doi:[10.1002/nme.2809](https://doi.org/10.1002/nme.2809)
62. Davenport WB (190) *Probability and random processes*. McGraw-Hill Book Company, New York
63. Madsen HO, Krenk S, Lind NC (1986) *Methods of structural safety*. Prentice Hall, Englewood Cliffs

# Chapter 8

## Stochastic Algebraic Equations

### 8.1 Introduction

Let  $A$  and  $B$  be random matrices defined on the probability spaces  $(\Omega_1, \mathcal{F}_1, P_1)$  and  $(\Omega_2, \mathcal{F}_2, P_2)$ , respectively that may or may not be distinct. Let  $U$  be an  $\mathbb{R}^d$ -valued random variable define by

$$AU = B, \quad (8.1)$$

referred to as a stochastic algebraic equation (SAE). The solution  $U$  of this equation is defined on the product probability space  $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P_1 \times P_2)$ . If  $A$  and  $B$  are defined on the same probability space  $(\Omega, \mathcal{F}, P)$ , so is  $U$ . SAEs may result from equilibrium conditions for physical systems with uncertain properties described by models with a finite number of degrees of freedom. They also result from time-invariant stochastic partial differential equations by discretizing both the physical space and the probability space. The physical space can be discretized by solving finite difference or finite element representations of these equations that have a finite number of degrees of freedom. The probability space can be discretized by replacing the random fields in the definition of stochastic differential equations by parametric models, that is, deterministic functions of spatial coordinates that depend on a finite number of random variables.

Monte Carlo simulation is the only general method for estimating the probability law of  $U$  in (8.1), but can be impractical in realistic applications since it involves repeated solutions of (8.1) for samples of  $A$  and  $B$ . This limitation has promoted the development of a broad range of approximate solutions for (8.1), that we divide into two classes depending on the degree of uncertainty in the random entries of  $A$  and  $B$ . The Monte Carlo simulation, stochastic reduced order models, stochastic Galerkin, stochastic collocation, and reliability methods are used in Sect. 8.2 to solve SAEs with random parameters of arbitrary uncertainty. The Taylor, perturbation, Neumann series, and equivalent linearization methods in Sect. 8.3 are for SAEs with random entries of relatively low uncertainty.

The implementation of most methods for solving (8.1) requires that  $A^{-1}$  exists a.s. Unfortunately, there is no simple criterion establishing whether a random matrix can be inverted with probability 1. Available results are limited to special classes of random matrices, for example, deterministic matrices with entries polluted by independent Gaussian variables [12] and other types of matrices [20], that are not sufficiently general to be useful in applications. For solutions of SAEs that require the a.s. existence of  $A^{-1}$ , we may need to construct histograms of the determinant of  $A$  or of the conditional number  $\|A\|\|A^{-1}\|$  of  $A$  to assess in an approximate manner whether or not  $A$  can be inverted, unless the structure of  $A$  suggests alternative techniques for assessing the existence of its inverse.

*Example 8.1.* Consider (8.1) with  $d = 3$  and

$$A = \begin{bmatrix} X_1 + X_2 & -X_2 & 0 \\ -X_2 & X_2 + X_3 & -X_3 \\ 0 & -X_3 & X_3 \end{bmatrix}$$

where  $\{X_i\}$  are uniformly distributed in  $(a_i, b_i)$ ,  $i = 1, 2, 3$ , with  $a_1 = 2.5$ ,  $b_1 = 3.5$ ,  $a_2 = 2$ ,  $b_2 = 3$ ,  $a_3 = 1.5$ , and  $b_3 = 2.5$ . The support of a histogram for the determinant of  $A$  obtained from 1000 independent samples of  $A$  is  $[8.05, 25.42]$ , and changes slightly if the sample size is increased. This suggests that  $A$  is invertible a.s., so that it is expected that (8.1) has a unique solution with probability 1.  $\diamond$

The special case of (8.1) with  $A$  deterministic is not discussed since, if  $A^{-1}$  exists, then  $U = A^{-1}B$  is a linear mapping of random vector  $B$ , so that the mean and correlation matrices of the solution are  $E[U] = A^{-1}E[B]$  and  $E[UU'] = A^{-1}E[BB'](A^{-1})'$ , respectively. Higher order statistics of  $U$  can be obtained efficiently by, for example, Monte Carlo simulation since the mapping  $B \mapsto U = A^{-1}B$  is known. If  $B$  is Gaussian, its first two moments suffice to find the probability law of  $U$ .

Also, interval solutions for (8.1) are not considered since probability statements based on these solutions are rather limited. For example, suppose the vectors  $V_a$  and  $V_b$  contain the random entries of  $A$  and  $B$  and that they take values in some bounded subsets  $D_a$  and  $D_b$ . Interval analysis constructs a set  $I$  that contains the solution  $U$  of (8.1) for  $V_a \in D_a$  and  $V_b \in D_b$ . If  $V_a$  and  $V_b$  are independent, then  $P(U \in I) \geq P(A \in D_a)P(B \in D_b)$  [7].

## 8.2 SAEs with Arbitrary Uncertainty

Monte Carlo simulation, stochastic reduced order models, stochastic Galerkin, stochastic collocation, and reliability methods are used to solve (8.1) approximately. Conditional analysis and state augmentation applied in the previous chapter to solve ordinary differential equations with random coefficients are of limited value when dealing with SAEs, and are not discussed.

The Monte Carlo, stochastic reduced order models, and stochastic collocation methods are non-intrusive, that is, existing deterministic software can be used to construct estimates and/or approximations for the solution of (8.1). On the other hand the stochastic Galerkin method does not have this feature, and is said to be intrusive. A non-intrusive Galerkin method has been proposed in [18].

### 8.2.1 General Considerations

Strong and weak solutions can be constructed for (8.1). Strong solutions exists if  $A^{-1}$  is defined a.s., and their existence is required by, for example, the Monte Carlo method. Weak solutions are particularly needed for the stochastic Galerkin method. We define weak solutions and assess their accuracy by bounds. Two types of bounds are presented. The first type is given by Theorem 8.1 and relates to weak solutions for (8.1). The second type are bounds on the discrepancy between solutions of distinct deterministic algebraic equations (Theorems 8.2, 8.3).

Let  $A$  be an  $(d, d)$ -random matrix defined on a probability space  $(\Omega, \mathcal{F}, P)$  and  $\mathcal{B} : L^2(\Omega, \mathcal{F}, P) \times L^2(\Omega, \mathcal{F}, P) \rightarrow \mathbb{C}$  a functional defined by

$$\mathcal{B}(U, V) = E[(AU)'V^*], \quad U, V \in L^2(\Omega, \mathcal{F}, P). \quad (8.2)$$

It is assumed that  $A$  is Hermitian and positive definite a.s. and that there exists constants  $c > 0$  and  $\alpha > 0$  such that  $|\mathcal{B}(U, V)| = |E[(AU)'V^*]| \leq c \|U\| \|V\|$  and  $\mathcal{B}(U, U) = E[(AU)'U^*] \geq \alpha \|U\|^2$ , where  $\|\cdot\|$  denotes the norm in  $L^2(\Omega, \mathcal{F}, P)$ . We say that  $\mathcal{B}$  with these properties is bounded and positive definite (Sect. B.4.2).

Note that  $(z^*)'Az = z'A^*z^* = z'A'z^* = (Az)'z^* \in \mathbb{R}$  for  $z \in \mathbb{C}^d$  since  $A$  is Hermitian, that is,  $A_{ij} = A_{ji}^*$ ,  $i, j = 1, \dots, d$ , and that  $(z^*)'Az > 0$  a.s. for  $z \in \mathbb{C}^d \setminus \{0\}$  since  $A$  is positive definite. Since  $A$  is Hermitian and positive definite, it admits the Cholesky decomposition  $A = L(L^*)'$ , where  $L$  is a lower triangular matrix.

**Definition 8.1** The weak solution of (8.1) is an  $\mathbb{R}^d$ -valued random variable  $U \in L^2(\Omega, \mathcal{F}, P)$  satisfying the equation

$$\mathcal{B}(U, V) = \langle B, V \rangle, \quad \forall V \in L^2(\Omega, \mathcal{F}, P), \quad (8.3)$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product in  $\mathbb{R}^d$ .

The weak form (8.3) of (8.1) results by multiplying  $AU = B$  with an arbitrary vector  $V \in L^2(\Omega, \mathcal{F}, P)$  and taking the expectation of the resulting equation. If  $B \in L^2(\Omega, \mathcal{F}, P)$ , the functional  $\langle B, V \rangle$  of  $V$  will be linear and bounded. The bilinear functional  $\mathcal{B}$  in (8.2) is bounded and elliptic under the stated assumptions. Various conditions can be imposed on matrix  $A$  such that  $\mathcal{B}$  is bounded and elliptic. A condition on  $A$  assuring that  $\mathcal{B}$  is bounded is given by the following example.

**Example 8.2** If there exists a constant  $\beta > 0$  such that  $\sum_{i=1}^d \left( \sum_{j=1}^d |A_{ij}| \right)^2 \leq \beta^2$  a.s., then the bilinear form  $\mathcal{B}$  is bounded, that is,  $|\mathcal{B}(U, V)| \leq \beta \|U\| \|V\|$  for all  $U, V \in L^2(\Omega, \mathcal{F}, P)$ .  $\diamond$

*Proof* We have  $|E[(AU)'V^*]| \leq (E[(AU)'(AU)^*])^{1/2} \|V\|$  by the Cauchy–Schwarz inequality, and

$$\begin{aligned} E[(AU)'(AU)^*] &= E \left[ \sum_{i=1}^d \left| \sum_{j=1}^d A_{ij} U_j \right|^2 \right] \leq E \left[ \sum_{i=1}^d \left( \sum_{j=1}^d |A_{ij}| |U_j| \right)^2 \right] \\ &\leq E \left[ \sum_{i=1}^d \left( \sum_{j=1}^d |A_{ij}| \|U\|_{\mathbb{R}^d} \right)^2 \right] = E \left[ \sum_{i=1}^d \left( \sum_{j=1}^d |A_{ij}| \right)^2 \|U\|_{\mathbb{R}^d}^2 \right] \\ &\leq \beta^2 E[\|U\|_{\mathbb{R}^d}^2] = \beta^2 \|U\|^2, \end{aligned}$$

where  $\|U\|_{\mathbb{R}^d}^2 = \sum_{k=1}^d |U_k|^2$  is the square of the Euclidean norm in  $\mathbb{R}^d$ . These observations give  $|\mathcal{B}(U, V)| = |E[(AU)'V^*]| \leq \beta \|U\| \|V\|$ .  $\blacktriangle$

Let  $X$  denote an  $\mathbb{R}^{d_x}$ -valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$  that collects all distinct random entries in matrices  $A$  and  $B$ . Let  $\tilde{X}$  be another  $d_x$ -dimensional random vector defined on the same probability space as  $X$ . Denote by  $\tilde{U}$  the solution of the SAE  $\tilde{A}\tilde{U} = \tilde{B}$ , where  $(\tilde{A}, \tilde{B})$  are  $(A, B)$  with  $\tilde{X}$  in place of  $X$ . If  $\tilde{X}(\Omega) \subseteq \Gamma = X(\Omega)$ , then  $\tilde{A}$  has the same properties as  $A$ , the functional  $\tilde{\mathcal{B}}(U, V) = E[(\tilde{A}U)'V^*]$  is well defined and has the same properties as  $\mathcal{B}$ .

We consider both strong and weak solutions for the SAEs  $AU = B$  and  $\tilde{A}\tilde{U} = \tilde{B}$ . The weak solutions of these equations satisfy  $\mathcal{B}(U, V) = E[B'V^*]$  and  $\tilde{\mathcal{B}}(U, V) = E[\tilde{B}'V^*]$  for all  $V \in L^2(\Omega, \mathcal{F}, P)$ . The relationship between weak and strong solutions follows from that between weak and strong convergence for random sequences in Hilbert spaces (Sect. B.4.3). The existence and uniqueness of weak solutions is guaranteed by the Lax–Milgram theorem (Theorem B.44) and Theorem 8.8 provided  $E[B'V^*]$  and  $E[\tilde{B}'V^*]$  are bounded for all  $V \in L^2(\Omega, \mathcal{F}, P)$ .

**Theorem 8.1** If  $U$  and  $\tilde{U}$  are weak solutions of  $AU = B$  and  $\tilde{A}\tilde{U} = \tilde{B}$ , matrices  $A$  and  $\tilde{A}$  have the properties in Definition 8.1,  $B, \tilde{B} \in L^2(\Omega, \mathcal{F}, P)$ , and  $\sum_{i=1}^d \left( \sum_{j=1}^d |\tilde{A}_{ij} - A_{ij}| \right)^2 \leq \mu(A, \tilde{A})^2$  a.s. for a constant  $\mu(A, \tilde{A}) > 0$ , then

$$\|U - \tilde{U}\| \leq \frac{1}{\alpha} (\|B - \tilde{B}\| + \mu(A, \tilde{A}) \|\tilde{U}\|), \quad (8.4)$$

where  $\alpha > 0$  is a constant such that  $\mathcal{B}(V, V) \geq \alpha \|V\|^2$  for all  $V \in L^2(\Omega, \mathcal{F}, P)$ .

*Proof* We have

$$\begin{aligned} |\mathcal{B}(U - \tilde{U}, V)| &\leq |\mathcal{B}(U, V) - \tilde{\mathcal{B}}(\tilde{U}, V)| + |\tilde{\mathcal{B}}(\tilde{U}, V) - \mathcal{B}(\tilde{U}, V)| \\ &= |E[(B - \tilde{B})' V^*]| + |E[(\tilde{A} - A)\tilde{U}]' V^*| \leq (\|B - \tilde{B}\| + \|(\tilde{A} - A)\tilde{U}\|) \|V\| \\ &\leq (\|B - \tilde{B}\| + \mu(A, \tilde{A})\|\tilde{U}\|) \|V\| \end{aligned}$$

by the Cauchy–Schwarz inequality, the assumption on the discrepancy between  $\tilde{A}$  and  $A$ , and the inequality

$$\|(\tilde{A} - A)\tilde{U}\|^2 = E[(\tilde{A} - A)\tilde{U}]' ((\tilde{A} - A)\tilde{U})^* \leq \mu(\tilde{A}, A)^2 \|\tilde{U}\|^2,$$

that follows by arguments similar to those used in Example 8.2. The inequalities  $|\mathcal{B}(U - \tilde{U}, V)| \leq (\|B - \tilde{B}\| + \mu(A, \tilde{A})\|\tilde{U}\|) \|V\|$  and  $\alpha \|V\|^2 \leq |\mathcal{B}(V, V)|$  for  $V = U - \tilde{U}$  give

$$\alpha \|U - \tilde{U}\|^2 \leq |\mathcal{B}(U - \tilde{U}, U - \tilde{U})| \leq (\|B - \tilde{B}\| + \mu(A, \tilde{A})\|\tilde{U}\|) \|U - \tilde{U}\|,$$

which yields (8.4) by division with  $\|U - \tilde{U}\|$ . As expected, the bound on the discrepancy between  $U$  and  $\tilde{U}$  depends on differences between  $(A, B)$  and  $(\tilde{A}, \tilde{B})$ , and vanishes as  $(\tilde{A}, \tilde{B})$  approaches  $(A, B)$ .  $\blacktriangle$

**Theorem 8.2** Let  $u^{(i)}$  and  $\tilde{u}^{(k)}$  denote solutions of (8.1) for  $X$  set equal to samples  $x^{(i)}$  and  $\tilde{x}^{(k)}$  of  $X$  and  $\tilde{X}$ , respectively. Denote by  $(A^{(i)}, B^{(i)})$  and  $(\tilde{A}^{(k)}, \tilde{B}^{(k)})$  the matrices  $(A, B)$  for  $X = x^{(i)}$  and  $X = \tilde{x}^{(k)}$ , respectively. If  $\|B^{(i)}\| \neq 0$ , then

$$\frac{\|\Delta u^{(k,i)}\|}{\|u^{(i)}\|} \leq \|(\tilde{A}^{(k)})^{-1}\| \|A^{(i)}\| \left( \frac{\|\Delta B^{(k,i)}\|}{\|B^{(i)}\|} + \frac{\|\Delta A^{(k,i)}\|}{\|A^{(i)}\|} \right), \quad (8.5)$$

where  $\Delta A^{(k,i)} = \tilde{A}^{(k)} - A^{(i)}$ ,  $\Delta B^{(k,i)} = \tilde{B}^{(k)} - B^{(i)}$ , and  $\Delta u^{(k,i)} = \tilde{u}^{(k)} - u^{(i)}$  and the norms are Euclidean norms for vectors and corresponding induced norms for matrices.

*Proof* Set  $(\tilde{A}^{(k)})^{-1} = (A^{(i)})^{-1} + \Delta^{-1}$ . Since  $\Delta u^{(k,i)} = (\tilde{A}^{(k)})^{-1} \tilde{B}^{(k)} - (A^{(i)})^{-1} B^{(i)}$ , we have

$$\begin{aligned} \Delta u^{(k,i)} &= ((A^{(i)})^{-1} + \Delta^{-1}) (B^{(i)} + \Delta B^{(k,i)}) - (A^{(i)})^{-1} B^{(i)} \\ &= (A^{(i)})^{-1} B^{(i)} + \Delta^{-1} B^{(i)} + (\tilde{A}^{(k)})^{-1} \Delta B^{(k,i)} - (A^{(i)})^{-1} B^{(i)} \\ &= (\tilde{A}^{(k)})^{-1} (\Delta B^{(k,i)} - \Delta A^{(k,i)} u^{(i)}). \end{aligned}$$

The expression of  $\Delta u^{(k,i)}$  results from  $I = \tilde{A}^{(k)}(\tilde{A}^{(k)})^{-1} = I + \Delta A^{(k,i)}(A^{(i)})^{-1} + \tilde{A}^{(k)}\Delta^{-1}$  or  $\tilde{A}^{(k)}\Delta^{-1} = -\Delta A^{(k,i)}(A^{(i)})^{-1}$ , which gives  $\Delta^{-1} = -(\tilde{A}^{(k)})^{-1}\Delta A^{(k,i)}(A^{(i)})^{-1}$  and  $\Delta^{-1}B^{(i)} = -(\tilde{A}^{(k)})^{-1}\Delta A^{(k,i)}u^{(i)}$ . The above inequality implies

$$\|\Delta u^{(k,i)}\| \leq \|(\tilde{A}^{(k)})^{-1}\| (\|\Delta B^{(k,i)}\| + \|\Delta A^{(k,i)}\| \|u^{(i)}\|),$$

which together with

$$\|\Delta B^{(k,i)}\| = \frac{\|\Delta B^{(k,i)}\|}{\|B^{(i)}\|} \|B^{(i)}\| \leq \frac{\|\Delta B^{(k,i)}\|}{\|B^{(i)}\|} \|A^{(i)}\| \|u^{(i)}\|,$$

yields (8.5).  $\blacktriangle$

**Theorem 8.3** *With the notation in Theorem 8.2, let  $\lambda_{\max}^{(i)}$  denote the largest real part of the eigenvalues of  $(A^{(i)} + (A^{(i)})')/2$ . If  $\lambda_{\max}^{(i)} < 0$  and the real part of the eigenvalues of  $\tilde{A}^{(k)}$  are negative, then*

$$\begin{aligned} \|\Delta u^{(k,i)}\| &\leq \frac{-1}{\lambda_{\max}^{(i)}} \|\Delta A^{(k,i)} \tilde{U}^{(k)} + \Delta B^{(k,i)}\| \\ &\leq \frac{-1}{\lambda_{\max}^{(i)}} \left( \|\Delta A^{(k,i)}\| \|(\tilde{A}^{(k)})^{-1}\| \|\tilde{B}^{(k)}\| + \|\Delta B^{(k,i)}\| \right). \end{aligned} \quad (8.6)$$

*Proof* Note that the condition  $\lambda_{\max}^{(i)} < 0$  is stronger than the requirement that the eigenvalues of  $A^{(i)}$  have strictly negative real parts [25] and that (8.6) does not hold for  $\lambda_{\max}^{(i)} \geq 0$ .

The discrepancy  $\|\tilde{Y}^{(k)}(t) - Y^{(i)}(t)\|$  between the solutions of the linear differential equations  $\dot{Y}^{(i)}(t) = A^{(i)} Y^{(i)}(t) - B^{(i)}$  and  $\dot{\tilde{Y}}^{(k)}(t) = \tilde{A}^{(k)} \tilde{Y}^{(k)}(t) - \tilde{B}^{(k)}$  can be bounded by ([13], Theorem 10.6, and Sect. 7.5.3.1 in this book)

$$\|\tilde{Y}^{(k)}(t) - Y^{(i)}(t)\| \leq e^{\lambda_{\max}^{(i)} t} \int_0^t e^{-\lambda_{\max}^{(i)} s} \|(A^{(i)} - \tilde{A}^{(k)}) \tilde{Y}^{(k)}(s) + (\tilde{B}^{(k)} - B^{(i)})\| ds. \quad (8.7)$$

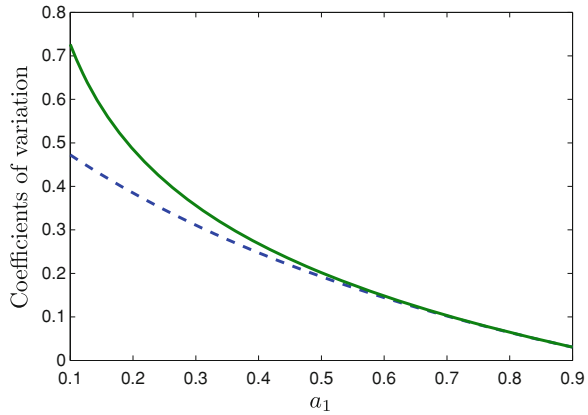
Since the above differential equations admit steady-state solutions by assumption,  $\lim_{t \rightarrow \infty} Y^{(i)}(t) = (A^{(i)})^{-1} B^{(i)}$  and  $\lim_{t \rightarrow \infty} \tilde{Y}^{(k)}(t) = (\tilde{A}^{(k)})^{-1} \tilde{B}^{(k)}$ , that is, the solutions of the algebraic equations  $A^{(i)} u^{(i)} = B^{(i)}$  and  $\tilde{A}^{(k)} \tilde{u}^{(k)} = \tilde{B}^{(k)}$ , respectively. The bound in (8.6) results from (8.7) in the limit as  $t \rightarrow \infty$ .  $\blacktriangle$

## 8.2.2 Monte Carlo Method

Estimates of moments and other properties of the solution  $U$  of (8.1) can be constructed from samples of this random vector calculated from samples of  $A$  and  $B$ , that are derived from samples of  $X$ . The methods in Sects. 2.13.1 to 2.13.3 can be used to generate samples of  $X$  and construct estimators for properties of  $U$ . It is assumed throughout this section that  $A^{-1}$  exists with probability 1.

*Example 8.3* Consider the one-dimensional problem  $AU = 1$  with  $A$  uniformly distributed in  $(a_1, a_2)$ ,  $0 < a_1 < a_2 < \infty$ . The unique solution of this stochastic algebraic equation is  $U = 1/A$ . Figure 8.1 shows with solid and dotted lines the coefficients of variation of  $U$  and  $A$  as a function of  $a_1$  for  $a_2 = 1$ . Since the

**Fig. 8.1** Coefficients of variation of  $A$  (dotted line) and  $U$  (solid line) for  $a_2 = 1$



uncertainty in  $U$  is much larger than that in  $A$  for small values of  $a_1$ , estimates of  $U$  need to be based on larger samples than those of  $A$ . However, it is not possible to determine the sample size needed to estimate moments and other properties of  $U$  to a specified accuracy since, in contrast to  $A$ , the probability law of  $U$  is unknown prior to calculations.  $\diamond$

*Proof* The mean and variance of  $A$  are  $(a_1 + a_2)/2$  and  $(a_2 - a_1)^2/12$  so that its coefficient of variation is  $(a_2 - a_1)/((a_1 + a_2)\sqrt{3})$ . The moments  $E[U^q] = (a_2^{1-q} - a_1^{1-q})/(1-q)/(a_2 - a_1)$  for  $q > 1$  and  $E[U^q] = \ln(a_2/a_1)/(a_2 - a_1)$  for  $q = 1$  can be used to calculate the coefficient of variation of  $U$ .  $\blacktriangle$

*Example 8.4* Let  $\hat{U}^q = \sum_{k=1}^n U_k^q/n$  be an estimator for the moment  $E[U^q]$  of order  $q$  for  $U$  in Example 8.3, where  $\{U_k\}$  are independent copies of  $U$ . The mean and variance of  $\hat{U}^q$  are  $E[U^q]$  and  $\text{Var}[U^q]/n$  so that  $\hat{U}^q$  is an unbiased estimator of  $E[U^q]$  whose accuracy improves with the sample size  $n$ .  $\diamond$

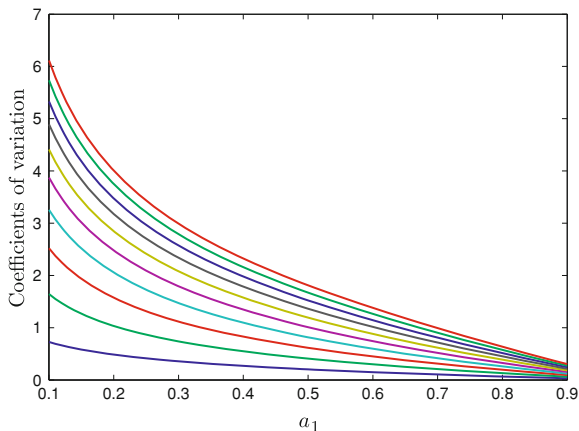
*Proof* We have  $E[\hat{U}^q] = \sum_{k=1}^n E[U_k^q]/n = E[U^q]$  so that  $\hat{U}^q$  is an unbiased estimator for  $E[U^q]$ . Since  $U_k$  are iid random variables, we have  $E[(\hat{U}^q)^2] = [\sum_{k=1}^n E[U_k^{2q}] + \sum_{k,l=1, k \neq l}^n E[U_k^q U_l^q]]/n^2 = [nE[U^{2q}] + (n^2 - n)E[U^q]^2]/n^2$ , which gives the stated result since  $E[\hat{U}^q] = E[U^q]$ .  $\blacktriangle$

Considerations in Examples 8.3 and 8.4 extend directly to the matrix equation given by (8.1). Let  $\{U_i^{(k)}\}$  be independent copies of  $U_i = \sum_{s=1}^d A_{is}^{-1} B_s$ , where  $A_{is}^{-1}$  denotes entry  $(i, s)$  of  $A^{-1}$ . The mean and variance of the estimator

$$\hat{U}_i^q = \frac{1}{n} \sum_{k=1}^n (U_i^{(k)})^q \quad (8.8)$$



**Fig. 8.2** Coefficients of variation of  $U^{2q}$  for  $q = 1, \dots, 10$  and  $a_2 = 1$



of  $E[U_i^q]$  are  $E[U_i^q]$  and  $\text{Var}[U_i^q]/n$ , that is,  $\hat{U}_i^q$  is unbiased estimator and its variance approaches 0 as  $n \rightarrow \infty$ . Similar unbiased estimators can be develop for the distributions of coordinates and functions of  $U$ . For example,

$$\hat{F}_h(\xi) = \frac{1}{n} \sum_{k=1}^n 1(h(U^{(k)}) \leq \xi), \quad (8.9)$$

is an estimator for the distribution  $F_h$  of the real-valued random variable  $h(U)$ , where  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  is a measurable function and  $\{U^{(k)}\}$  are independent copies of  $U$ .

The computational effort for constructing estimators of the types in (8.8) and (8.9) has two distinct components, the time for calculating a single sample of the solution  $U$  of (8.1) and the sample size  $n$  needed to estimate accurately properties of functions of  $U$ . Relatively large computation times needed to obtained a single sample of  $U$  can be an insufficient argument for labeling the Monte Carlo inefficient. The label may be inadequate if, for example, estimators with satisfactory accuracy can be obtained from a relatively small number of samples, in which case Monte Carlo can be competitive. For example, Fig. 8.2 shows coefficients of variation for  $U^{2q}$ ,  $q = 1, \dots, 10$ , as a function of  $a_1$ , where  $U$  is the solution of the stochastic algebraic equation  $AU = 1$  in Example 8.3 for  $a_2 = 1$ . The ordinates of the graphs increase with  $q$  for each  $a_1$ . The plots suggest that large samples of  $U$  are needed to obtain satisfactory estimates for  $E[U^{2q}]$  if  $a_1$  is close to 0. On the other hand,  $E[U^{2q}]$  can be estimated accurately from a relatively small number samples of  $U$  if  $a_1$  is away from 0 and, for example,  $q \leq 2$ , that is, moments of  $U$  up to order 4.

### 8.2.3 Stochastic Reduced Order Model Method

Let  $X$  be an  $\mathbb{R}^d$ -valued random variable including all random parameters in the definition of the stochastic algebraic equation  $AU = B$  given by (8.1). The idea is to approximate  $U$  by the solution  $\tilde{U}$  of a version of (8.1) obtained by replacing  $X$  with

a stochastic reduced order model (SROM)  $\tilde{X}$  of it, that is, a simple random vector taking  $m \geq 1$  distinct values  $(\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)})$  in the range  $\Gamma = X(\Omega)$  of  $X$  with probabilities  $(p_1, \dots, p_m)$ . The defining parameters  $(\tilde{x}^{(k)}, p_k)$ ,  $k = 1, \dots, m$ , of  $\tilde{X}$  are such that the probability laws of  $X$  and  $\tilde{X}$  are similar in some sense (Sect. A.3). Bounds on statistics of  $\|U - \tilde{U}\|$  are developed and used to measure the performance of  $\tilde{U}$ . It is assumed as in the previous section that  $A$  is invertible a.s.

Two SROM-based methods for solving SAEs are discussed. The first method approximates the solution  $U$  of a SAE by a SROM  $\tilde{U}$  with samples  $\{\tilde{u}_k\}$  obtained by solving deterministic versions of (8.1) with samples  $\{\tilde{x}_k\}$  of a SROM  $\tilde{X}$  of  $X$  in place of  $X$ . The samples  $\{\tilde{u}_k\}$  have the same probabilities  $\{p_k\}$  as the samples of  $\tilde{X}$ . The second method approximates the mapping  $X \mapsto U$  by a piecewise linear function given by hyperplanes tangent to it at  $\{\tilde{u}_k\}$ . The resulting representation for  $X \mapsto U$  is more accurate than that in the first method, but its implementation is less simple. In addition to  $\{u_k\}$ , we need to find the gradients of  $U$  with respect to the coordinates of  $X$  at  $\{\tilde{u}_k\}$ , construct a Voronoi tessellation with centers  $\{\tilde{x}_k\}$  in the range  $\Gamma = X(\Omega)$  of  $X$ , and estimate statistics of  $U$  from a piecewise linear representation of mapping  $X \mapsto U$  by Monte Carlo simulation. We refer to the first and second methods as SROM- and extended stochastic reduced order model (ESROM)-based solutions (Sects. A.3 and A.4).

### 8.2.3.1 Stochastic Reduced Order Models (SROMs)

The construction of SROM-based solutions for (8.1) involves three steps. First, a SROM  $\tilde{X}$  of  $X$  needs to be constructed. Optimization algorithms with objective functions measuring the discrepancy between properties of  $X$  and  $\tilde{X}$  under the constraints  $p_k \geq 0$ ,  $k = 1, \dots, m$ , and  $\sum_{k=1}^m p_k = 1$  can be used to find  $(\tilde{x}^{(k)}, p_k)$ ,  $k = 1, \dots, m$ , (Sect. A.3). The defining parameters of  $\tilde{X}$  depend solely on the probability law of  $X$  and their determination does not involve solutions of (8.1).

Second,  $m$  solutions  $\tilde{u}^{(k)}$  of deterministic versions of (8.1) with  $X$  set equal to  $\tilde{x}^{(k)}$ ,  $k = 1, \dots, m$ , need to be calculated. These solutions and the probabilities of  $\tilde{X}$  define a SROM  $\tilde{U}$  for  $U$ . The defining parameters of  $\tilde{U}$  are  $(\tilde{u}^{(k)}, p_k)$ ,  $k = 1, \dots, m$ .

Third,  $U$  is approximated by  $\tilde{U}$ . Properties of  $\tilde{U}$  can be obtained by elementary calculations. For example, the distribution  $P(\tilde{U}_j \leq z)$  and the moment of order  $r \geq 1$  of coordinate  $\tilde{U}_j$ ,  $j = 1, \dots, d$ , of  $\tilde{U}$  are  $P(\tilde{U}_j \leq z) = \sum_{k=1}^m 1(\tilde{u}_j^{(k)} \leq z) p_k$  and  $E[\tilde{U}_j^r] = \sum_{k=1}^m (\tilde{u}_j^{(k)})^r p_k$ , respectively.

We have seen that the construction of a SROM  $\tilde{U}$  for  $U$  involves solutions of deterministic versions of (8.1) with  $X$  set equal to points in  $\Gamma = X(\Omega)$ . The implementation of the Monte Carlo simulation and stochastic collocation methods also involve solutions of deterministic versions of (8.1). The points in  $\Gamma$  used by both Monte Carlo and SROM-based solutions depend on the probability law of  $X$ . On the other hand, stochastic collocation solutions use points in  $\Gamma$  that are unrelated to the probability law of  $X$  (Sect. 8.2.5).

*Example 8.5* Consider the stochastic algebraic equation  $AU = 1$  with  $A = X \sim U(a, b)$ . Let  $\tilde{x}^{(1)} < \dots < \tilde{x}^{(m)}$  be  $m$  equally spaced points in  $(a, b)$  with  $\tilde{x}^{(1)} - a = b - \tilde{x}^{(m)} = \varepsilon > 0$ . To complete the definition of  $\tilde{X}$ , we need to specify the probabilities  $(p_1, \dots, p_m)$  of the samples of  $\tilde{X}$ . The accuracy of SROM solutions of (8.1) depends on the defining parameters  $(\tilde{x}^{(k)}, p_k)$ ,  $k = 1, \dots, m$ , of  $\tilde{X}$ . For example, the errors in the mean and the standard deviation of  $\tilde{U}$  are  $-42.27\%$  and  $-81.78\%$  for  $(a, b) = (0.01, 3)$ ,  $\varepsilon = 0.2$  and  $m = 4$ . They decrease to  $2.76\%$  and  $-15.89\%$ , respectively, for  $\varepsilon = 0.03$  and  $m = 20$ . The errors of the mean and the standard deviation of  $\tilde{U}$  are  $0.0385\%$  and  $-1.2980\%$  for  $(a, b) = (1, 3)$ ,  $\varepsilon = 0.2$ , and  $m = 4$ .  $\diamond$

*Proof* The probabilities  $(p_1, \dots, p_m)$  of the range  $(\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)})$  of  $\tilde{X}$  have been selected to minimize the objective function

$$e = \alpha_1 \int (F(x) - \tilde{F}(x))^2 dx + \alpha_2 \sum_{r=1}^{\tilde{r}} (E[X^r] - E[\tilde{X}^r])^2, \quad (8.10)$$

under the constraints  $p_k \geq 0$ ,  $k = 1, \dots, m$ , and  $\sum_{k=1}^m p_k = 1$ , where  $\alpha_1, \alpha_2 > 0$  are constants,  $F(x) = P(X \leq x) = (x - a)/(b - a)$ ,  $x \in (a, b)$ ,  $\tilde{F}(x) = P(\tilde{X} \leq x) = \sum_{k=1}^m p_k 1(\tilde{x}^{(k)} \leq x)$ ,  $E[X^r] = (b^{r+1} - a^{r+1})/(b - a)/(r + 1)$ , and  $E[\tilde{X}^r] = \sum_{k=1}^m (\tilde{x}^{(k)})^r p_k$ . The distribution and the moments of the SROM  $\tilde{U}$  of  $U$  are  $P(\tilde{U} \leq u) = \sum_{k=1}^m p_k 1(u \leq \tilde{u}^{(k)})$  and  $E[\tilde{U}^r] = \sum_{k=1}^m (\tilde{u}^{(k)})^r p_k$ , where  $\tilde{u}^{(k)}$  is the solution of  $AU = 1$  for  $X = \tilde{x}^{(k)}$ . Numerical results in this example are for  $\alpha_1 = \alpha_2 = 1$ .  $\blacktriangle$

The stochastic algebraic equation in Example 8.5 is also solved in Examples 8.15 and 8.16 by the stochastic Galerkin and collocation methods. The accuracy of the approximate means and standard deviations by all these methods is remarkable for  $(a = 1, b = 3)$  but varies from method to method for  $(a = 0.01, b = 3)$ .

The bounds in Theorems 8.2 and 8.3 developed for deterministic algebraic equations corresponding to samples of  $X$  and  $\tilde{X}$  can be extended to stochastic algebraic equations. Let  $\{x^{(i)}, i = 1, \dots, n\}$  be a collection of independent samples of  $X$  that is sufficiently large to provide an accurate characterization for the probability law of  $X$ . Let  $\{\mathcal{C}_k, k = 1, \dots, m\}$ ,  $m \ll n$ , be a partition of  $(x^{(1)}, \dots, x^{(n)})$  such that  $p_k = n_k/n$ , where  $n_k$  denotes the cardinality of  $\mathcal{C}_k$ . Let  $\zeta$  be a measurable function mapping the members of  $\mathcal{C}_k$  into  $\tilde{x}^{(k)}$ , that is,  $\zeta(x^{(i)}) = \tilde{x}^{(k)}$  for  $x^{(i)} \in \mathcal{C}_k$ . The mapping  $\zeta$  can be constructed by an algorithm in Sect. A.3 that is also outlined here for convenience.

Suppose  $X$  is described satisfactorily by  $n$  independent samples  $(x^{(1)}, \dots, x^{(n)})$ . This set of samples can be partitioned by the following two-step procedure. In the first step, we construct a partition  $\mathcal{C}'_k$  by assigning  $x^{(i)}$  to  $\mathcal{C}'_k$  if it is closer to  $\tilde{x}^{(k)}$  than any other  $\tilde{x}^{(l)}$ ,  $l \neq k$ , that is,  $x^{(i)}$  is assigned to  $\mathcal{C}'_k$  if  $d(x^{(i)}, \tilde{x}^{(k)}) < d(x^{(i)}, \tilde{x}^{(l)})$ ,  $l \neq k$ , where  $d$  is a metric in  $\Gamma = X(\Omega)$ . If  $d(x^{(i)}, \tilde{x}^{(k)}) = d(x^{(i)}, \tilde{x}^{(l)})$  for two distinct indices  $(k, l)$ , then  $x^{(i)}$  is assigned to either  $\mathcal{C}'_k$  or  $\mathcal{C}'_l$ . Generally, the cardinalities  $n'_k$  of the resulting clusters  $\mathcal{C}'_k$  do not satisfy the condition  $p_k = n'_k/n$ . In the second step, we eliminate the members of the clusters  $\mathcal{C}'_k$  with  $n'_k/n > p_k$  that are the

farthest from  $\tilde{x}^{(k)}$  till the reduced versions  $\mathcal{C}_k''$  of these clusters satisfy the condition  $n_k''/n \simeq p_k$ , where  $n_k''$  denotes the cardinality of  $\mathcal{C}_k''$ . The members extracted from the clusters  $\mathcal{C}_k'$  with  $n_k'/n > p_k$  are assigned to the clusters  $\mathcal{C}_k'$  with  $n_k'/n < p_k$  based on their closeness to the nuclei  $\tilde{x}^{(k)}$  of these clusters and the requirement  $n_k''/n \simeq p_k$ . The algorithm delivers a partition  $\{\mathcal{C}_k, k = 1, \dots, m\}$  of  $(x^{(1)}, \dots, x^{(n)})$  such that the members  $x^{(i)}$  of  $\mathcal{C}_k$  are mapped into  $\tilde{x}^{(k)}$  and the probability that  $X$  takes values in  $\mathcal{C}_k$  is  $p_k \simeq n_k/n$ .

With these considerations the bounds in (8.5) and (8.6) extend directly to stochastic algebraic equations. For example, the expectation of the discrepancy  $\|\tilde{U} - U\|$  between the exact and the approximate solutions of (8.1) can be bounded by

$$\begin{aligned} E[\|\tilde{U} - U\|] &\leq \frac{1}{n} \sum_{k=1}^m \sum_{x^{(i)} \in \mathcal{C}_k} \frac{-1}{\lambda_{\max}^{(i)}} \|\Delta A^{(k,i)} + \Delta B^{(k,i)}\| \\ &\leq \frac{1}{n} \sum_{k=1}^m \sum_{x^{(i)} \in \mathcal{C}_k} \frac{-1}{\lambda_{\max}^{(i)}} \left( \|\Delta A^{(k,i)}\| \|(\tilde{A}^{(k)})^{-1}\| \|\tilde{B}^{(k)}\| + \|\Delta B^{(k,i)}\| \right) \end{aligned} \quad (8.11)$$

by (8.6). Similar bounds can be constructed for other statistics of the discrepancy between  $U$  and  $\tilde{U}$ .

The bounds in Theorems 8.1, 8.2, and 8.3 on the discrepancy between properties of  $U$  and  $\tilde{U}$  show that the approximate solutions  $\tilde{U}$  converge to the exact solution  $U$  of (8.1) as the discrepancy between  $(\tilde{A}, \tilde{B})$  and  $(A, B)$  vanishes. However, their calculation is impractical, as illustrated by the bound on  $E[\|\tilde{U} - U\|]$  in (8.11).

**Theorem 8.4** *Let  $g : G \rightarrow \mathbb{R}$  be a differentiable function such that  $\|\nabla g(\xi)\| \leq M$ , where  $M > 0$  is a constant. If the mapping  $X \mapsto U$  defined by (8.1) is Lipschitz, that is, solutions  $u', u''$  of (8.1) with  $X$  equal to  $x', x'' \in \mathbb{R}^{d_x}$  are such that  $\|u' - u''\| \leq c\|x' - x''\|$  for a constant  $c > 0$ , the discrepancy between the probabilities  $P(g(U) \leq z)$  and  $P(g(\tilde{U}) \leq z)$  can be bounded by*

$$|P(g(U) \leq z) - P(g(\tilde{U}) \leq z)| \leq \frac{1}{n} \sum_{k=1}^m \sum_{x^{(i)} \in \mathcal{C}_k} \zeta^{(k,i)}(z), \quad (8.12)$$

where  $\zeta^{(k,i)}(z) = |1(g(u^{(i)}) \leq z) - 1(g(\tilde{u}^{(k)}) \leq z)| = |1cM\|x^{(i)} - \tilde{x}^{(k)}\| \text{sign}(z - g(\tilde{u}^{(k)})) \leq z - g(\tilde{u}^{(k)}) - 1(g(\tilde{u}^{(k)}) \leq z)|$ .

*Proof* As previously, let  $u^{(i)}$  and  $\tilde{u}^{(k)}$  be solutions of (8.1) for  $X = x^{(i)}, i = 1, \dots, n$ , and  $X = \tilde{x}^{(k)}, k = 1, \dots, m$ , respectively. For  $x^{(i)} \in \mathcal{C}_k$ , there exists  $\chi^{(k,i)}$  on the segment connecting  $u^{(i)}$  to  $\tilde{u}^{(k)}$  such that  $g(u^{(i)}) = g(\tilde{u}^{(k)}) + J$  with  $J = \nabla g(\chi^{(k,i)}) \cdot (u^{(i)} - \tilde{u}^{(k)})$  by the mean value theorem. Suppose  $g(\tilde{u}^{(k)}) \leq z$  so that  $1(g(\tilde{u}^{(k)}) \leq z) = 1$ . If  $J \leq z - g(\tilde{u}^{(k)})$ , then  $1(g(u^{(i)}) \leq z) = 1$  so that  $\zeta^{(k,i)}(z) = 0$ . Otherwise,  $1(g(u^{(i)}) \leq z) = 0$  and  $\zeta^{(k,i)}(z) = 1$ . We construct a lower bound on  $1(g(u^{(i)}) \leq z)$  which will yield an upper bound on  $\zeta^{(k,i)}(z)$ . Since  $|J| \leq \|\nabla g(\chi^{(k,i)})\| \|u^{(i)} - \tilde{u}^{(k)}\| \leq cM \|x^{(i)} - \tilde{x}^{(k)}\|$  implying  $J \leq |J| \leq cM \|x^{(i)} - \tilde{x}^{(k)}\|$ , we have  $1(g(u^{(i)}) \leq$

$z) \geq 1(cM\|x^{(i)} - \tilde{x}^{(k)}\| \leq z - g(\tilde{u}^{(k)}))$  so that  $\zeta^{(k,i)}(z) \leq |1(cM\|x^{(i)} - \tilde{x}^{(k)}\| \leq z - g(\tilde{u}^{(k)})) - 1(g(\tilde{u}^{(k)}) \leq z)|$ . Similar considerations for the case  $g(\tilde{u}^{(k)}) \geq z$  give  $1(g(u^{(i)}) \leq z) \leq 1(cM\|x^{(i)} - \tilde{x}^{(k)}\| \geq g(\tilde{u}^{(k)}) - z)$  (Exercise 8.3).  $\blacktriangle$

*Example 8.6* Under the conditions of Theorem 8.4 and the assumption that the random variables  $g(U)$  and  $g(\tilde{U})$  have finite expectations, we have

$$|E[g(U)] - E[g(\tilde{U})]| = |E[g(U) - g(\tilde{U})]| \leq \frac{cM}{n} \sum_{k=1}^m \sum_{x^{(i)} \in \mathcal{C}_k} \|x^{(i)} - \tilde{x}^{(k)}\|, \quad (8.13)$$

where the constants  $c, M > 0$  are as in Theorem 8.4.  $\diamond$

*Proof* Since  $g(u^{(i)}) - g(\tilde{u}^{(k)}) = \nabla g(\chi^{(k,i)}) \cdot (u^{(i)} - \tilde{u}^{(k)})$ , we have  $|g(u^{(i)}) - g(\tilde{u}^{(k)})| \leq \|\nabla g(\chi^{(k,i)})\| \|u^{(i)} - \tilde{u}^{(k)}\|$ , where  $\chi^{(k,i)}$  is on the segment with ends  $u^{(i)}$  and  $\tilde{u}^{(k)}$ . This bound on  $|g(u^{(i)}) - g(\tilde{u}^{(k)})|$  and  $|E[g(U) - g(\tilde{U})]| = (1/n) \sum_{k=1}^m \sum_{x^{(i)} \in \mathcal{C}_k} |g(u^{(i)}) - g(\tilde{u}^{(k)})|$  give (8.13).  $\blacktriangle$

*Example 8.7* Consider the stochastic algebraic equation (8.1) with  $d = 1$ ,  $A = X$  a random variable uniformly distributed in  $(a, b)$ , and  $B = 1$ . Three SROMs  $\tilde{X}$  with  $m = 10, 20$ , and  $40$  samples have been constructed for  $X$ . The range of these models is as in Example 8.5, that is,  $\tilde{x}^{(1)} = a + \varepsilon$ ,  $\tilde{x}^{(m)} = b - \varepsilon$ , and the rest of the samples of  $\tilde{X}$  are equally spaced in the range  $(\tilde{x}^{(1)}, \tilde{x}^{(m)})$  with  $\varepsilon = 0.1, 0.03$ , and  $0.018$  for  $m = 10, 20$ , and  $40$ , respectively. The scaled version of the upper bound in (8.13), that is,  $(1/n) \sum_{k=1}^m \sum_{x^{(i)} \in \mathcal{C}_k} \|x^{(i)} - \tilde{x}^{(k)}\|$ , is equal to  $0.0944, 0.0644$ , and  $0.0585$  for  $n = 1000$  and  $m = 10, 20$ , and  $40$ , respectively. The cardinalities of the largest cluster  $\mathcal{C}_k$  are  $107, 54$ , and  $26$  for  $m = 10, 20$ , and  $40$ .  $\diamond$

The bounds in (8.12) and (8.13) do not account for statistical uncertainty since their construction assumes that  $X$  is a simple  $\mathbb{R}^{d_x}$ -valued random variable with equally likely samples  $(x^{(1)}, \dots, x^{(n)})$ . These bounds can be modified to account for the fact that the estimated properties of  $X$  depend on the particular samples  $(x^{(1)}, \dots, x^{(n)})$  considered in analysis and the sample size  $n$ . For example, let  $\hat{I}_P(z) = (1/n) \sum_{i=1}^n 1(g(U^{(i)}) \leq z)$  and  $\hat{I}_U = (1/n) \sum_{i=1}^n g(U^{(i)})$  be estimators of  $P(g(U) \leq z)$  and  $E[g(U)]$ , respectively, where  $U^{(i)}$  are independent copies of  $U$ . Since the mean and variance of, for example, the estimator  $\hat{I}_U = (1/n) \sum_{i=1}^n g(U^{(i)})$  are  $E[g(U)]$  and  $\text{Var}[g(U)]/n$ , respectively, we conclude that  $\hat{I}_U$  is an unbiased estimator for  $E[g(U)]$  with the property  $P(|\hat{I}_U - E[g(U)]| > \varepsilon) \leq \text{Var}[g(U)]/(n\varepsilon^2)$  that holds for  $\varepsilon > 0$  arbitrary by Chebyshev's inequality.

The contribution of statistical uncertainty to the previous bounds on the discrepancy between  $\tilde{U}$  and  $U$  can be incorporated simply. We have

$$\begin{aligned} |P(g(U) \leq z) - P(g(\tilde{U}) \leq z)| &\leq |P(g(U) \leq z) - \hat{I}_P(z)| + |\hat{I}_P(z) - P(g(\tilde{U}) \leq z)| \\ |E[g(U)] - E[g(\tilde{U})]| &\leq |E[g(U)] - \hat{I}_U| + |\hat{I}_U - E[g(\tilde{U})]|, \end{aligned} \quad (8.14)$$

where the terms  $|\hat{I}_P(z) - P(g(\tilde{U}) \leq z)|$  and  $|\hat{I}_U - E[g(\tilde{U})]|$  are those in (8.12) and (8.13) while the terms  $|P(g(U) \leq z) - \hat{I}_P(z)|$  and  $|E[g(U)] - \hat{I}_U|$  relate to

statistical uncertainty. The latter terms are defined as in (8.8) and (8.9), and vanish as  $n \rightarrow \infty$ .

It is possible to replace the distribution  $\tilde{F}(u) = \sum_{k=1}^m p_k 1(\tilde{u}^{(k)} \leq u)$  of  $\tilde{U}$  by a smooth version of it. The substitution principle suggests to smooth  $\tilde{F}(u)$  by convolving it with a continuous distribution  $G$  with probability mass concentrated at the origin of  $\mathbb{R}^d$  ([16], Sect. 4.5). For  $d = 1$  the smooth version of  $\tilde{F}$  has the form

$$\tilde{F}_{\text{smooth}}(u) = \int G(u - v) d\tilde{F}(y) = \sum_{k=1}^m p_k G(u - u^{(k)}). \quad (8.15)$$

If  $G$  is concentrated at 0, the discrepancy  $|\tilde{F}(u) - \tilde{F}_{\text{smooth}}(u)|$  between the distributions  $\tilde{F}$  and  $\tilde{F}_{\text{smooth}}$  is small. If  $G$  is differentiable,  $\tilde{F}_{\text{smooth}}$  has a density  $\tilde{f}_{\text{smooth}}(u) = \sum_{k=1}^m p_k g(u - u^{(k)})$ , where  $g$  denotes the derivative of  $G$ .

We conclude this section with two examples. The first example solves (8.1) with

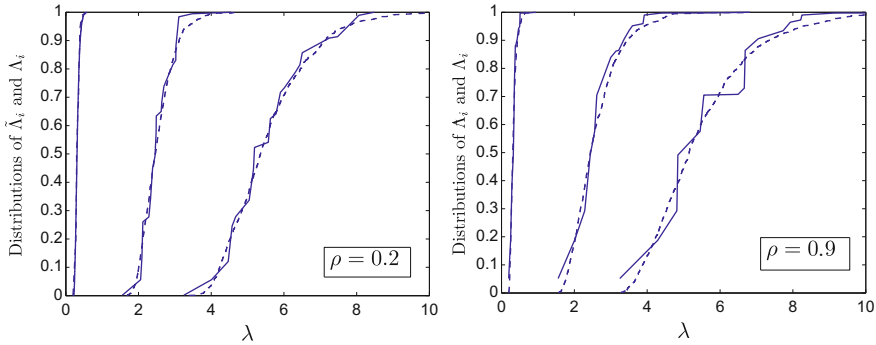
$$A = \begin{bmatrix} X_1 + X_2 - X_2 & 0 \\ -X_2 & X_2 + X_3 - X_3 \\ 0 & -X_3 & X_3 \end{bmatrix} \quad (8.16)$$

by SROMs, where  $X_i = F^{-1} \circ \Phi(G_i)$ ,  $F$  is a Gamma distribution with shift, shape, and decay parameters  $a = 1$ ,  $k = 2$ , and  $\eta = 3$ , and  $\{G_i\}$  are standard Gaussian variables with  $E[G_i G_j] = \rho^{|i-j|}$ ,  $i, j = 1, 2, 3$ . The right side  $B$  of (8.1) is the unit vector in  $\mathbb{R}^3$ . The second example solves the eigenvalue problem  $AU = \Lambda U$ .

*Example 8.8* Let  $U \in \mathbb{R}^3$  be the solution of (8.1) with  $A$  in (8.16) and  $B$  the unit vector in  $\mathbb{R}^3$ . A SROM  $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \tilde{X}_3)$  of  $X = (X_1, X_2, X_3)$  with dimension  $m = 10$  has been constructed. The first three columns and the last column of matrix  $\alpha$  in (8.17) give the samples  $(\tilde{x}_1^{(k)}, \tilde{x}_2^{(k)}, \tilde{x}_3^{(k)})$ ,  $k = 1, \dots, m$ , of  $\tilde{X}$  and the probabilities  $p_k$ ,  $k = 1, \dots, m$ , respectively, for  $\rho = 0.9$ . The first sample of  $\tilde{X}$  has been selected to coincide with the smallest values of  $X_j$ ,  $j = 1, 2, 3$ .

$$\alpha = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 & 0.0482 \\ 2.0110 & 1.8888 & 2.1573 & 0.0969 \\ 1.8569 & 1.5090 & 1.2753 & 0.0603 \\ 1.4923 & 1.3309 & 1.7060 & 0.1146 \\ 1.2891 & 1.4006 & 1.4669 & 0.2010 \\ 1.3987 & 1.1906 & 1.1748 & 0.0965 \\ 2.2620 & 2.1742 & 2.2356 & 0.0643 \\ 2.5874 & 2.7606 & 2.1229 & 0.1351 \\ 1.7373 & 1.7505 & 1.6375 & 0.1831 \\ 2.3473 & 2.1110 & 1.8373 & 0.0001 \end{bmatrix} \quad (8.17)$$

The samples  $(\tilde{u}^{(1)}, \dots, \tilde{u}^{(m)})$  of the SROM  $\tilde{U}$  have been obtained by solving ten deterministic problems (8.1) with  $X = \tilde{x}^{(k)}$ . Moments of any order  $r$  of the coordinates of  $\tilde{U}$  can be calculated simply from  $E[\tilde{U}_j^r] = \sum_{k=1}^{10} (u_j^{(k)})^r p_k$ . The moments



**Fig. 8.3** Distributions of the eigenvalues of  $A$  based on a SROM for  $\tilde{X}$  with  $m = 20$  (solid lines) and Monte Carlo simulation (dotted lines) for  $\rho = 0.2$  (left panel) and  $\rho = 0.9$  (right panel)

of the coordinates of  $\tilde{U}$  up to order 4 are in error by less than 8% relative to Monte Carlo estimates of these moments based on 1000 independent samples of  $X$ .  $\diamond$

*Example 8.9* Random eigenvalue problems can be solved by SROMs in a similar manner. First, a SROM  $\tilde{X}$  with parameters  $(\tilde{x}^{(k)}, p_k)$ ,  $k = 1, \dots, m$ , is constructed for  $X = (X_1, X_2, X_3)$ . Second, the eigenvalues  $\{\tilde{\lambda}_j^{(k)}\}$ ,  $j = 1, 2, 3$ , of  $A$  with  $X = \tilde{x}^{(k)}$ ,  $k = 1, \dots, m$ , are calculated, so that  $m$  distinct deterministic eigenvalue problems need to be solved. Third, SROMs are assembled for the random eigenvalues  $\{\tilde{\Lambda}_j\}$  of  $A$  from the eigenvalues  $\{\tilde{\lambda}_j^{(k)}\}$ ,  $j = 1, 2, 3$ , and their probabilities  $p_k$ ,  $k = 1, \dots, m$ .

The dotted lines in Fig. 8.3 are Monte Carlo estimates of the distributions of the eigenvalues  $\{\Lambda_j\}$  obtained from 1000 independent samples of  $A$ . The solid lines are the distributions of the eigenvalues  $\{\tilde{\Lambda}_j\}$  based on a SROM  $\tilde{X}$  with  $m = 20$  samples. The left and right panels in the figure are for correlation coefficients  $\rho = 0.2$  and  $\rho = 0.9$ , respectively. The maximum error of the first four moments of  $\{\tilde{\Lambda}_j\}$  with respect to corresponding Monte Carlo estimates based on 1000 independent samples are under 5.38% for  $m = 10$  and 3.97% for  $m = 20$  if  $\rho = 0.2$  and under 4.47% for  $m = 10$  and 4.11% for  $m = 20$  if  $\rho = 0.9$ .  $\diamond$

### 8.2.3.2 Extended Stochastic Reduced Order Models (ESROMs)

Technical details on extended stochastic reduced order models (ESROMs) can be found in Sect. A.4. The ESROM-based solution for SAEs is based on the piecewise linear representation

$$U_L(X) = \sum_{k=1}^m [\tilde{u}_k + \nabla \tilde{u}_k \cdot (X - \tilde{x}_k)] 1(X \in \Gamma_k) \quad (8.18)$$

of mapping  $X \mapsto U$ , where  $\{\tilde{u}_k\}$  are solutions of (8.1) for  $X$  replaced with the samples  $\{\tilde{x}_k\}$  of a SROM  $\tilde{X}$  of  $X$ ,  $\{\nabla \tilde{u}_k = (\partial U / \partial x_1, \dots, \partial U / \partial x_{d_x})\}$  denote gradients of  $U$  at  $\{X = \tilde{x}_k\}$ , and  $\Gamma_k$  are the cells of a Voronoi tessellation in  $\Gamma = X(\Omega)$  with centers  $\{\tilde{x}_k\}$ . Note that  $\{A_k = X^{-1}(\Gamma_k)\}$  is a measurable partition of  $\Omega$ .

Properties of  $U_L$  can be calculated simply from its expression by Monte Carlo simulation. For example the moment of order  $q \geq 1$  and the distribution of a coordinate  $U_{L,j}$ ,  $j = 1, \dots, d$ , of  $U_L$  can be estimated from

$$\begin{aligned} E[U_{L,j}^q] &\simeq \sum_{k=1}^m \frac{n_k}{n} \left[ \frac{1}{n_k} \sum_{x_i \in \Gamma_k} (\tilde{u}_{k,j} + w_{ki,j})^q \right] \quad \text{and} \\ P(U_{L,j} \leq u) &\simeq \sum_{k=1}^m \frac{n_k}{n} \left[ \frac{1}{n_k} \sum_{x_i \in \Gamma_k} 1(\tilde{u}_k + w_{ki,j} \leq u) \right], \end{aligned} \quad (8.19)$$

where  $\tilde{u}_{k,j}$  and  $w_{ki,j}$  denote the  $j$ th coordinate of  $\tilde{u}_k$  and  $\nabla \tilde{u}_k \cdot (x_i - \tilde{x}_k)$ ,  $\{x_i\}$  are  $n$  independent samples of  $X$ , and  $n_k$  denotes the number of samples  $\{x_i\}$  in  $\Gamma_k$ . The estimates in (8.19) follow from properties of conditional expectation. For  $h : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$  measurable, we have  $E[h(U_L(X))] = E\{E[h(U_L(X)) \mid \mathcal{G}]\}$  and  $E[h(U_L(X)) \mid \mathcal{G}] = \sum_{k=1}^m [(1/P(A_k)) \int_{A_k} h(U_L(X)) dP] 1_{A_k}$ , where  $\{A_k = X^{-1}(\Gamma_k)\}$  is a measurable partition of  $\Omega$  and  $\mathcal{G} = \sigma(A_1, \dots, A_m)$  (Sect. 2.8). Hence,  $E[h(U_L(X))]$  can be viewed as a sum of local averages  $\{(1/P(A_k)) \int_{A_k} h(U_L(X)) dP\}$  weighted by probabilities  $\{P(A_k)\}$ . In (8.19), these local averages and probabilities are estimated by  $\{(1/n_k) \sum_{x_i \in \Gamma_k} h(U_L(x_i))\}$  and  $\{n_k/n\}$ , respectively.

In the remainder of this section, we construct bounds on the discrepancy between first and higher order approximations,  $U_L(X)$  and  $U_q(X)$ , of  $U(X)$  and  $U(X)$ , and solve a random eigenvalue problem and a SAE by ESROMs. The bounds result from properties of the Taylor series. It is assumed that  $U(X)$  is a real-valued random variable. The extension to vector-valued solutions  $U(X)$  poses no conceptual difficulty.

**Theorem 8.5** *If  $d_x = 1$ ,  $U(X)$  has  $q \geq 1$  continuous derivatives in  $\Gamma$  a.s.,  $U^{(q+1)}$  exists in  $\Gamma$  a.s., and  $X$  has finite moments of order  $q$ , then*

$$\begin{aligned} E[|U(X) - \tilde{U}_q(X)|] &\leq \sum_{k=1}^m P(A_k) \frac{M_{q,k}}{(q+1)!} E[|X - \tilde{x}_k|^{q+1} \mid A_k] \\ &\leq \frac{M_q}{(q+1)!} \sum_{k=1}^m P(A_k) E[|X - \tilde{x}_k|^{q+1} \mid A_k] \end{aligned} \quad (8.20)$$

where

$$\tilde{U}_q(X) = \sum_{k=1}^m 1(X \in \Gamma_k) \left[ \sum_{r=0}^q \frac{U^{(r)}(\tilde{x}_k)}{r!} (X - \tilde{x}_k)^r \right], \quad (8.21)$$

$|U^{(q+1)}(x)| \leq M_{q,k}$  a.s. in  $\Gamma_k$ ,  $0 < M_{q,k} < \infty$ , and  $M_q \geq M_{q,k}$  for  $k = 1, \dots, m$ .



*Proof* The Taylor expansion of order  $q$  of  $U(x)$  in  $\Gamma_k$  about  $\tilde{x}_k$  is  $U(x) = \tilde{U}_q(x) + \sum_{k=1}^m 1(x \in \Gamma_k) R_q(c_k, x)$ , where  $R_q(c_k, x) = [U^{(q+1)}(c_k)/(q+1)!](x - \tilde{x}_k)^{q+1}$  is the remainder and  $c_k \in \Gamma_k$  ([4], Theorem 21.1). We have

$$\begin{aligned} E[|U(X) - \tilde{U}_q(X)|] &= \sum_{k=1}^m P(A_k) E\left[\frac{U^{(q+1)}(c_k)}{(q+1)!} (X - \tilde{x}_k)^{q+1} \mid A_k\right] \\ &\leq \sum_{k=1}^m P(A_k) \frac{M_{q,k}}{(q+1)!} E[|X - \tilde{x}_k|^{q+1} \mid A_k], \end{aligned}$$

which yields (8.20). The terms  $P(A_k) E[|X - \tilde{x}_k|^{q+1} \mid A_k] = \int_{\Gamma_k} |x - \tilde{x}_k|^{q+1} dF(x)$  in (8.20) can be calculated by numerical integration or Monte Carlo simulation, where  $F$  denotes the distribution of  $X$ .  $\blacktriangle$

*Example 8.10* The bound in (8.20) can be used to assess the accuracy of, for example, the piecewise linear and quadratic approximations  $\tilde{U}_L(X) = \tilde{U}_1(X)$  and  $\tilde{U}_Q(X) = \tilde{U}_2(X)$ . These bounds for the mapping  $X \mapsto U = 1.25 - (X - 0.5)^2$ ,  $X$  a real-valued Beta random variable with range  $\Gamma = X(\Omega) = [a, b]$ ,  $\{\Gamma_k = [(k-1)\Delta x, k\Delta x]\}$ ,  $\Delta x = (b-a)/m$ , shape parameters  $(\zeta, \eta)$ ,  $a = 1/2$ ,  $b = 4$ ,  $\zeta = 1/2$ , and  $\eta = 3$  are

$$\begin{aligned} E[|U(X) - \tilde{U}_L(X)|] &\leq \frac{M_1}{2!} \sum_{k=1}^m \int_{\Gamma_k} |x - \tilde{x}_k|^2 dF(x) \text{ and} \\ E[|U(X) - \tilde{U}_Q(X)|] &\leq \frac{M_2}{3!} \sum_{k=1}^m \int_{\Gamma_k} |x - \tilde{x}_k|^3 dF(x), \end{aligned}$$

where  $F$  denotes the distribution of  $X$ ,  $M_1 = 2$ , and  $M_2 = 0$ . The discrepancy  $E[|U(X) - \tilde{U}_Q(X)|]$  is zero since the mapping  $X \mapsto U(X)$  is quadratic.  $\diamond$

**Theorem 8.6** *If  $d > 1$ ,  $U(X)$  has continuous partial derivatives of order  $q > 1$  in  $\Gamma$  a.s., the partial derivatives of order  $q+1$  exists in  $\Gamma$  a.s., and  $X$  has finite moments of order  $q$ , then*

$$E[|U(X) - \tilde{U}_q(X)|] \leq \frac{d}{q+d} \sum_{k=1}^m M_{q,k} P(A_k) E\left[\frac{\|X - \tilde{x}_k\|^{q+d}}{\|X - c_k\|^{d-1}} \mid A_k\right] \sum^{(q)} \prod_{j=1}^d \frac{1}{q_j!} \quad (8.22)$$

where

$$\tilde{U}_q(X) = \sum_{k=1}^m 1(X \in \Gamma_k) \left[ \sum^{(\leq q)} \left( \prod_{j=1}^d \frac{(x_j - \tilde{x}_{k,j})^{q_j}}{q_j!} \right) \frac{\partial^{q_1+\dots+q_d} U(\tilde{x}_k)}{\partial x_1^{q_1} \dots \partial x_d^{q_d}} \right], \quad (8.23)$$

$\sum^{(q)} = \sum_{q_1, \dots, q_d \geq 0; q_1 + \dots + q_d = q}$ ,  $c_k$  is a point on the open interval with ends  $\tilde{x}_k$  and  $x$ ,  $\sum^{(\leq q)} = \sum_{q_1, \dots, q_d \geq 0; q_1 + \dots + q_d \leq q}$ ,  $|\partial^{q+1} U(x) / (\partial x_1^{q_1} \dots \partial x_i^{q_i+1} \dots \partial x_d^{q_d})| \leq$

$M_{q,k}$  for all  $x \in \Gamma_k$ ,  $0 < M_{q,k} < \infty$ , and  $\{x_j\}$  and  $\{\tilde{x}_{k,j}\}$  are the coordinates of  $x$  and  $\tilde{x}_k$ .

*Proof* The multivariate Taylor formula at  $x \in \mathbb{R}^d$  at expansion point  $\tilde{x}_k$  has the form  $U(x) = \tilde{U}_q(x) + \sum_{k=1}^m 1(x \in \Gamma_k) R_q(c_k, x)$ , where

$$\begin{aligned} R_q(c_k, x) &= \frac{\sum_{i=1}^d \alpha_i \|c_k - \tilde{x}_k\| \|x - \tilde{x}_k\|^n \left[ \sum^{(q)} \left( \prod_{j=1}^d \frac{\alpha_j^{q_j}}{q_j!} \right) \frac{\partial^{q_1+\dots+q_d} U(c_k)}{\partial x_1^{q_1} \dots \partial x_i^{q_i+1} \dots \partial x_d^{q_d}} \right]}{\frac{1}{N(d)} \sum_{i=1}^d (q_i + 1) \frac{\|c_k - \tilde{x}_k\|}{\|x - \tilde{x}_k\|} \sum^{(q)} \left( \frac{\|x - c_k\|}{\|x - \tilde{x}_k\|} \right)^{q+d-1}} \\ &= \frac{\|x - \tilde{x}_k\|^{q+d}}{(q+d)\|x - c_k\|^{d-1}} \sum_{i=1}^d \alpha_i \sum^{(q)} \left( \prod_{j=1}^d \frac{\alpha_j^{q_j}}{q_j!} \right) \frac{\partial^{q+1} U(c_k)}{\partial x_1^{q_1} \dots \partial x_i^{q_i+1} \dots \partial x_d^{q_d}} \end{aligned}$$

denotes the remainder for the Taylor approximation of order  $q$ ,  $\{\alpha_j\}$  are the cosines of direction  $x - \tilde{x}_k$ , and  $N(d)$  gives the number of terms in  $\sum^{(q)}$ . For  $x \in \Gamma_k$ , the remainder can be bounded by

$$|R_q(c_k, x)| \leq \frac{M_{q,k} d}{q+d} \frac{\|x - \tilde{x}_k\|^{q+d}}{\|x - c_k\|^{d-1}} \sum^{(q)} \prod_{j=1}^d \frac{1}{q_j!}$$

since  $|\alpha_i| \leq 1$ . This bound on the local error of the approximation  $\tilde{U}_q$  of order  $q$  of  $U$  and  $E[|U(X) - \tilde{U}_q(X)|] = \sum_{k=1}^m P(A_k) E[R_q(c_k, X) | A_k]$  yield (8.22).  $\blacktriangle$

For the special case  $d_x = 1$ , the bound in (8.22) takes the form

$$\begin{aligned} E[|U(X) - \tilde{U}_q(X)|] &\leq \frac{1}{q+1} \sum_{k=1}^m M_{q,k} P(A_k) E[|X - \tilde{x}_k|^{q+1} | A_k] \frac{1}{q!} \\ &\leq \frac{M_q}{(q+1)!} \sum_{k=1}^m P(A_k) E[|X - \tilde{x}_k|^{q+1} | A_k], \end{aligned}$$

with  $M_q \geq M_{q,k}$ , and coincides with the bound in (8.20).

**Theorem 8.7** *The discrepancy  $E[|U(X) - \tilde{U}_q(X)|]$  can be made as small as desired by refining the partition  $\{\Gamma_k\}$  of  $\Gamma$ .*

*Proof* Consider a sequence of partitions  $\{\Gamma_k\}$  of  $\Gamma$  whose diameter decreases with  $m$ , that is,  $\max_{1 \leq k \leq m, x', x'' \in \Gamma_k} \|x' - x''\| \sim O(\varepsilon_m)$ ,  $\varepsilon_m > 0$ , and  $\varepsilon_m \rightarrow 0$  as  $m \rightarrow \infty$ . Since  $P(A_k) E[\|X - \tilde{x}_k\|^{q+d} / \|X - c_k\|^{d-1} | A_k] \sim O(\varepsilon_m^{q+1})$ , we have  $E[|U(X) - \tilde{U}_q(X)|] \rightarrow 0$  as  $m \rightarrow \infty$  by (8.20).  $\blacktriangle$

Similar results hold for functions  $h(U)$  of  $U$  under some conditions, for example, functions  $h : \mathbb{R} \rightarrow \mathbb{R}$  that are measurable and Lipschitz continuous. Results in Theorems 8.5–8.7 extend to vector-valued solutions  $U$ .

We conclude the section with two examples involving solutions of a random eigenvalue problem  $AU = \Lambda U$  and a stochastic algebraic equation  $AU = B$ . The random matrix  $A$  in both problems is

$$A = \begin{bmatrix} X_1 + X_2 & -X_2 & 0 \\ -X_2 & X_2 + X_3 & -X_3 \\ 0 & -X_3 & X_3 \end{bmatrix}, \quad (8.24)$$

where  $X_i = F^{-1} \circ \Phi(G_i)$ ,  $F$  is a Beta distribution with range  $[a, b]$  and shape parameters  $(\zeta, \eta)$ ,  $G_i \sim N(0, 1)$ ,  $E[G_i G_j] = \rho^{|i-j|}$ ,  $i, j = 1, 2, 3$ , and  $\rho \in (-1, 1)$ .

*Example 8.11* Let  $X = (X_1, X_2, X_3)$  be random vector collecting the random entries of matrix  $A$  in (8.24) and  $\tilde{X}$  a SROM for  $X$  with parameters  $\{\tilde{x}_k, p_k\}$ ,  $k = 1, \dots, m$ . The defining parameters of the corresponding SROMs for the eigenvalues  $\{\Lambda_i(X)\}$  of  $A$  are  $\{\tilde{\lambda}_{i,k} = \Lambda_i(\tilde{x}_k), p_k\}$ .

The piecewise linear approximations in (8.18) for the eigenvalues  $\Lambda_i(X)$  of  $A$  are

$$\Lambda_{L,i}(X) = \sum_{k=1}^m \left[ \tilde{\lambda}_{i,k} + \sum_{r=1}^d \tilde{\lambda}_{i,k}^{(r)} (X_r - \tilde{x}_{k,r}) \right] 1(X \in \Gamma_k), \quad (8.25)$$

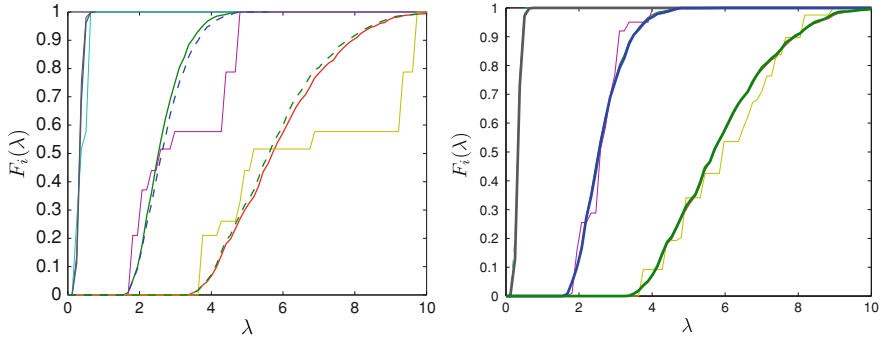
where  $\tilde{\lambda}_{i,k} = \Lambda_i(\tilde{x}_k)$ ,

$$\lambda_{i,k}^{(r)} = \frac{\partial \Lambda_i(X)}{\partial X_r} \Big|_{(X=\tilde{x}_k)} = -\frac{c_{1,k}^{(r)} \lambda_{i,k}^{n-1} + \dots + c_{n-1,k}^{(r)} \lambda_{i,k} + c_{n,k}^{(r)}}{n \lambda_{i,k}^{n-1} + (n-1) c_{1,k} \lambda_{i,k}^{n-2} + \dots + c_{n-1,k}}, \quad k = 1, \dots, m, \quad (8.26)$$

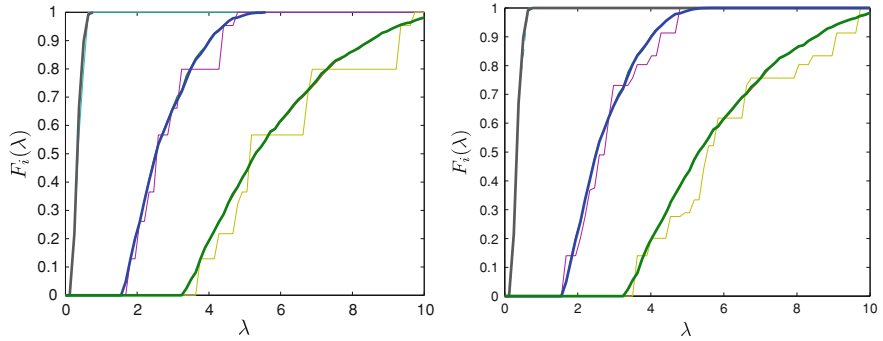
$\{c_{i,k}\}$  are the coefficients  $\{C_i\}$  of the characteristic equation  $\det(A - \Lambda I) = \Lambda^n + C_1 \Lambda^{n-1} + \dots + C_{n-1} \Lambda + C_n = 0$  for  $X = \tilde{x}_k$ ,  $c_{i,k}^{(r)} = \partial C_i(X) / \partial X_r$  at  $X = \tilde{x}_k$ ,  $\{X_r\}$  are the coordinates of  $X$ , and  $\{\tilde{x}_{k,r}\}$  denote the coordinates of  $\tilde{x}_k$  ([11], Sect. 8.3.2.3). The samples  $\{\tilde{x}_k\}$  of  $\tilde{X}$  are the centers of the Voronoi tessellation  $\{\Gamma_k\}$  in  $\Gamma$ . Similar approximations can be constructed for the eigenvectors of  $A$  ([11], Sect. 8.3.2.3).

The heavy solid lines in Figs. 8.4 and 8.5 are Monte Carlo estimates for the distributions  $F_i(\lambda) = P(\Lambda_i \leq \lambda)$ ,  $i = 1, 2, 3$ , of the eigenvalues of  $A$  based on 1000 independent samples of this matrix. The thin solid and heavy dotted lines are SROM- and ESROM-based solutions for the distributions  $F_i(\lambda)$ . The left and right panels in the figures are for  $m = 10$  and  $m = 20$ . The plots in Figs. 8.4 and 8.5 are for  $\rho = 0.2$  and  $\rho = 0.9$ . Some ESROM-based solutions are not visible since they coincide with Monte Carlo estimates of  $F_i(\Lambda)$  at the figure scale.

The accuracy of the SROM- and ESROM-based solutions improves significantly as the size of  $\tilde{X}$  increases from  $m = 10$  to  $m = 20$ , in agreement with Theorem 8.7. Superior approximations result for  $\rho = 0.9$  since a random vector with strongly correlated coordinates has less uncertainty in the sense that, given a coordinate of this vector, the uncertainty in its unspecified coordinates is reduced significantly. ESROM-based solutions are superior since they use representations of



**Fig. 8.4** Distributions  $F_i(\lambda)$  (heavy solid lines) and approximations of these distributions by SROM (thin solid lines) and ESROM (heavy dotted lines) for  $\rho = 0.2$ ,  $m = 10$  (left panel), and  $m = 20$  (right panel)



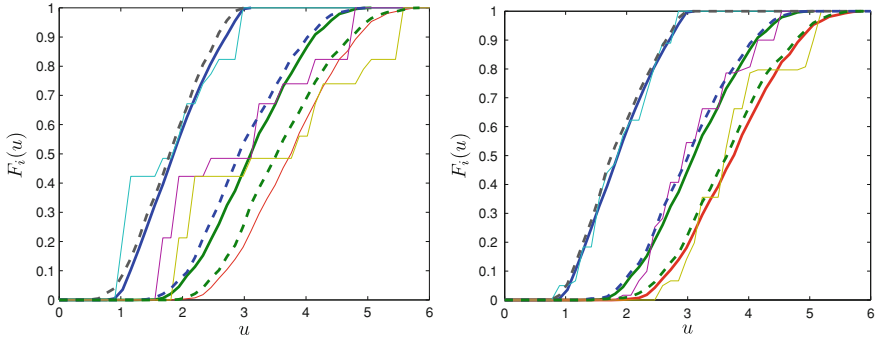
**Fig. 8.5** Distributions  $F_i(\lambda)$  (heavy solid lines) and approximations of these distributions by SROM (thin solid lines) and ESROM (heavy dotted lines) for  $\rho = 0.9$ ,  $m = 10$  (left panel), and  $m = 20$  (right panel)

mapping  $X \mapsto \Lambda$  that are more accurate than those for SROM-based solutions. Note also that SROMs  $\tilde{U}$  for  $U$  cannot provide any information on  $U$  outside their range  $[\min_{1 \leq k \leq m}(U(\tilde{z}_k)), \max_{1 \leq k \leq m}(U(\tilde{z}_k))]$ .  $\diamond$

**Example 8.12** Let  $U$  be the solution of  $AU = B$  with  $A$  in (8.24) and  $B$  the unit vector in  $\mathbb{R}^3$ . We use the SROMs  $\tilde{X}$  in Example 8.11 to construct  $\tilde{U}_L(X)$ . Let  $\{\tilde{x}_k, p_k\}$ ,  $k = 1, \dots, m$ , be the defining parameters of a SROM  $\tilde{X}$  of  $X$ , and denote by  $U(\tilde{x}_k)$  the solution of the deterministic algebraic equation  $AU = B$  for  $X = \tilde{x}_k$ ,  $k = 1, \dots, m$ . The gradients of  $U(X)$  at  $X = \tilde{x}_k$  satisfy the deterministic algebraic equations

$$A(X) \frac{\partial U(X)}{\partial X_r} = - \frac{\partial A(X)}{\partial X_r} U(X), \quad r = 1, 2, 3, \quad (8.27)$$

for  $X = \tilde{x}_k$ ,  $k = 1, \dots, m$ . Note that the deterministic algebraic equations for  $U(\tilde{x}_k)$  and  $\partial U(\tilde{x}_k)/\partial X_r$ ,  $r = 1, 2, 3$ , have the same operator and



**Fig. 8.6** Monte Carlo estimates for  $F_i(u)$  (heavy solid lines), SROM-based solutions (thin solid lines), and ESROM-based solutions (heavy dotted lines) for  $\rho = 0.2$ ,  $m = 10$  (left panel), and  $m = 20$  (right panel)

$$A(\tilde{x}_k) = \begin{bmatrix} \tilde{x}_{k,1} + \tilde{x}_{k,2} & -\tilde{x}_{k,2} & 0 \\ -\tilde{x}_{k,2} & \tilde{x}_{k,2} + \tilde{x}_{k,3} & -\tilde{x}_{k,3} \\ 0 & -\tilde{x}_{k,3} & \tilde{x}_{k,3} \end{bmatrix} \quad (8.28)$$

for each  $k = 1, \dots, m$ . The partial derivatives  $\{\partial A(X)/\partial X_r\}$  result simply from the definition of  $A$ , for example,

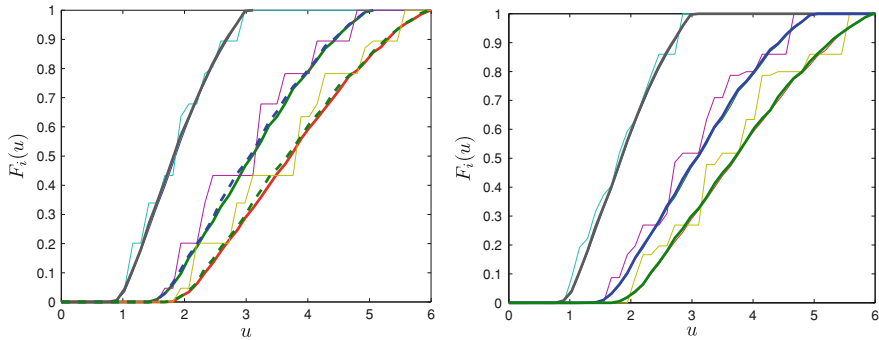
$$\frac{\partial A(\tilde{x}_k)}{\partial X_2} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The ESROM-based solution in (8.18) has the expression

$$\tilde{U}_L(X) = \sum_{k=1}^m \left[ U(\tilde{x}_k) - \sum_{r=1}^3 A(\tilde{x}_k)^{-1} \frac{\partial A(\tilde{x}_k)}{\partial X_r} U(\tilde{x}_k) (X_r - \tilde{x}_{k,r}) \right] 1(X \in \Gamma_k). \quad (8.29)$$

Since the expression of  $\tilde{U}_L(X)$  is known, its properties can be calculated efficiently and accurately by Monte Carlo simulation.

Figure 8.6 shows estimates of the distributions  $F_i(u) = P(U_i(X) \leq u)$ ,  $i = 1, 2, 3$ , of the coordinates of  $U(X)$  obtained from 1000 independent samples of  $A$  (heavy solid lines). SROM- and ESROM-based solutions of these distributions are in thin solid and heavy dotted lines for  $\rho = 0.2$ . Similar plots are in Fig. 8.7 for  $\rho = 0.9$ . The left and right panels are for  $m = 10$  and  $m = 20$ . The accuracy of both SROM- and ESROM-based solutions increases with  $m$  and  $\rho$ . As in the previous example, ESROM-based solutions are superior. The ESROM-based solutions in Fig. 8.7 for  $m = 20$  are indistinguishable from Monte Carlo estimates of  $F_i(u)$  at the figure scale.  $\diamond$



**Fig. 8.7** Monte Carlo estimates for  $F_i(u)$  (heavy solid lines), SROM-based solutions (thin solid lines), and ESROM-based solutions (heavy dotted lines) for  $\rho = 0.9, m = 10$  (left panel), and  $m = 20$  (right panel)

### 8.2.4 Stochastic Galerkin Method

Let  $X = (X_1, \dots, X_{d_x})$  be as previously an  $\mathbb{R}^{d_x}$ -valued random variable including the random parameters in matrices  $A$  and  $B$  of (8.1). There exists a nonlinear mapping,

$$\begin{aligned} F_1(X_1) &= V_1 \\ F_{2|1}(X_2 | X_1) &= V_2 \\ &\vdots \\ F_{d_x|d_x-1, \dots, 1}(X_{d_x} | X_{d_x-1}, \dots, X_1) &= V_{d_x}, \end{aligned} \quad (8.30)$$

relating  $X$  to an  $d_x$ -dimensional vector  $V = (V_1, \dots, V_{d_x})$  whose coordinates are independent random variables uniformly distributed in  $(0, 1)$  ([23], Theorem 3.5.1). The mapping depends on the distribution  $F_1$  of  $X_1$  and the distributions  $F_{j|j-1, \dots, 1}(x_j | x_{j-1}, \dots, x_1) = \int_{u \leq x_j} f_{j|j-1, \dots, 1}(u | x_{j-1}, \dots, x_1) du$ ,  $j = 2, \dots, d_x$ , where  $f_{j|j-1, \dots, 1}(\cdot | x_{j-1}, \dots, x_1)$  denotes the probability density function of the conditional random variable  $X_j | (X_{j-1} = x_{j-1}, \dots, X_1 = x_1)$ .

An alternative form of (8.30) can be obtained by using the representations  $V_j = \Phi(G_j)$ ,  $j = 1, \dots, d_x$ , where  $\{G_j\}$  are independent  $N(0, 1)$  variables. This version of (8.30) defines a mapping  $G = (G_1, \dots, G_{d_x}) \mapsto X = h(G)$  relating  $X$  to a  $d_x$ -dimensional vector  $G$  with independent  $N(0, 1)$  coordinates.

**Example 8.13** Let  $X$  be a bivariate translation vector defined by  $X_j = F_j^{-1}(\Phi(G_j))$ ,  $j = 1, 2$ , where  $F_j$ ,  $j = 1, 2$ , are continuous distributions,  $(G_1, G_2)$  are Gaussian variables with  $E[G_j] = 0$ ,  $E[G_j^2] = 1$ ,  $j = 1, 2$ , and  $E[G_1 G_2] = \rho$ ,  $|\rho| < 1$ . The mapping  $(V_1, V_2) \mapsto (X_1, X_2)$  in (8.30) is

$$\begin{aligned}
X_1 &= F_1^{-1}(V_1) \\
X_2 \mid X_1 &= F_2^{-1}\left(\Phi\left(\rho\Phi^{-1}(F_1(X_1)) + \sqrt{1-\rho^2}\Phi^{-1}(V_2)\right)\right). \quad (8.31)
\end{aligned}$$

This representation is not needed for translation vectors as considered in this example, since  $X = (X_1, X_2)$  is related simply to  $G = (G_1, G_2)$  in this case.  $\diamond$

*Proof* The formulas in (8.31) result from  $X_j = F_j^{-1}(\Phi(G_j))$ ,  $j = 1, 2$ , the property  $G_2 \mid G_1 \stackrel{d}{=} \rho G_1 + \sqrt{1-\rho^2}N(0, 1)$  of Gaussian variables, and the relationship between  $N(0, 1)$  and  $U(0, 1)$  variables.  $\blacktriangle$

The solution  $U$  of (8.1) with  $X = h(G)$  given by a version of (8.30) relating  $X$  to  $G$  can be viewed as an unknown function of  $G$ , where  $G = (G_1, \dots, G_{d_x})$  is a Gaussian vector with independent standard normal coordinates. Hence, the coordinates of both  $X$  and  $U$  can be represented by series of Hermite polynomials of  $G$ , referred to as polynomial chaos expansions. For example, the coordinate  $U_i$  of  $U$  admits the representation

$$\begin{aligned}
U_i &= \sum_{n=0}^{\infty} \sum_{n_1+n_2+\dots=n} a_{n_1, n_2, \dots}^i \prod_{k=1}^{d_x} \frac{1}{\sqrt{n_k!}} H_{n_k}(G_k) \\
&= \sum_{n=0}^{\infty} \sum_{n_1+n_2+\dots=n} a_{n_1, n_2, \dots}^i \psi_{n_1, n_2, \dots}(G), \quad i = 1, \dots, d, \quad (22)
\end{aligned} \quad (8.32)$$

where  $\psi_{n_1, n_2, \dots}(G) = \prod_k H_{n_k}(G_k) / \sqrt{n_k!}$ ,  $n_k \geq 0$  are integers,  $\{H_{n_k}\}$  denote Hermite polynomials,  $\{a_{n_1, n_2, \dots}^i\}$  are unknown deterministic coefficients, and  $\{G_k\}$  are the coordinates of  $G$ . The equality in (8.32) holds in the mean square sense ([10], Sect. 3.3.6, [17], Sect. 9.5, and Sect. B.6.1 of this book). Note that the coordinates of  $X$  admit representations as in (8.32) and that their coefficients can be calculated from the probability law of  $X$ , which is known.

For numerical calculations the infinite series in (8.32) needs to be truncated to a finite number of terms, so that the coordinates  $U_i$  of  $U$  are approximated by

$$\tilde{U}_i = \sum_{n=0}^{\bar{n}} \sum_{n_1+n_2+\dots=n} a_{n_1, n_2, \dots}^i \psi_{n_1, n_2, \dots}(G), \quad i = 1, \dots, d, \quad (8.33)$$

where  $\bar{n} \geq 1$  is an integer. The representation of  $U$  in (8.33) is used to construct the stochastic Galerkin solution for (8.1) ([10], Sect. 3.2.3). Similarly, the coordinates  $X_i$  of  $X$  can be approximated by finite sums as in (8.33), that are denoted by  $\tilde{X}_i$ . Note that the meaning of  $\tilde{X}$  and  $\tilde{U}$  in this section differs from that in the previous section.

The construction of stochastic Galerkin solutions is based on the following facts, that are discussed in Appendix B and summarized here for convenience. The space  $L^2(\Omega, \mathcal{F}, P)$  admits the decomposition  $L^2(\Omega, \mathcal{F}, P) = \oplus_{n=0}^{\infty} K_n$ , where the members of subspaces  $K_n$  are called homogeneous chaoses of order  $n$  (Theorem B.73

and Definition B.44). Hence, any member  $\phi$  of  $L^2(\Omega, \mathcal{F}, P)$  can be given in the form  $\phi = \sum_{n=0}^{\infty} \phi_n$ , where  $\phi_n \in K_n$  (Theorem B.73). The collection of functions  $\{\mathcal{H}_{n_1, n_2, \dots}, n_k \geq 0, n_1 + n_2 + \dots = n\}$  is an orthonormal basis for  $K_n$  (Theorem B.73), where  $\mathcal{H}_{n_1, n_2, \dots} = \prod_{k=1}^{\infty} H_{n_k}(G_k)/\sqrt{n_k!}$  and  $\{G_k\}$  are independent  $N(0, 1)$  variables. The functions  $\{\mathcal{H}_{n_1, n_2, \dots}, n_1 + n_2 + \dots = n, n = 0, 1, \dots\}$  define an orthonormal basis for the Hilbert space  $L^2(\Omega, \mathcal{F}^B, P)$  so that every  $\phi \in L^2(\Omega, \mathcal{F}^B, P)$  admits the representation  $\phi = \sum_{n=0}^{\infty} \sum_{n_1 + n_2 + \dots = n} a_{n_1, n_2, \dots} \mathcal{H}_{n_1, n_2, \dots}$ , where  $n_1, n_2, \dots \geq 0$  and  $a_{n_1, n_2, \dots} = E[\phi \mathcal{H}_{n_1, n_2, \dots}]$  (Theorem B.75).

The stochastic Galerkin method assumes that both the random elements of (8.1) and the solution of this equation can be represented satisfactorily by members of a subspace  $\mathcal{W}_{\tilde{n}} = \bigoplus_{n=0}^{\tilde{n}} K_n$  of  $L^2(\Omega, \mathcal{F}, P)$ . For example,  $U_i$  in (8.33) is a member of  $\mathcal{W}_{\tilde{n}} = \bigoplus_{n=0}^{\tilde{n}} K_n$ . The notations in this section are commonly used in applications, and differ slightly from those in Appendix B.

The stochastic Galerkin method gives weak solutions for approximate versions of  $AU = B$  in (8.1) obtained by representing the random elements in the definition of this equation by members of  $\mathcal{W}_{\tilde{n}}$ . Let  $(\tilde{A}, \tilde{B})$  be the matrices  $(A, B)$  with  $\tilde{X}$  in place of  $X$ , where  $\tilde{X}$  denotes the projection of  $X$  on  $\mathcal{W}_{\tilde{n}}$ . Conditions for the existence and uniqueness of the weak solution of  $\tilde{A}\tilde{U} = \tilde{B}$  can be established by arguments similar to those in [1] used to construct weak solutions for stochastic partial differential equation.

**Definition 8.2** The weak solution  $\tilde{U} \in \mathcal{W}_{\tilde{n}}$  of  $\tilde{A}\tilde{U} = \tilde{B}$  is given by

$$\tilde{\mathcal{B}}(\tilde{U}, W) = E[B'W^*], \quad \forall W \in \mathcal{W}_{\tilde{n}}, \quad (8.34)$$

where  $\tilde{\mathcal{B}}(\tilde{U}, W) = E[(\tilde{A}\tilde{U})'W^*]$ .

**Theorem 8.8** Let  $(\tilde{A}, \tilde{B})$  be  $(A, B)$  with  $\tilde{X}$  in place of  $X$ . If  $\tilde{\mathcal{B}}$  is a bounded, positive definite bilinear form, the weak form

$$\tilde{\mathcal{B}}(\tilde{U}, W) = E[\tilde{B}'W^*], \quad \forall W \in \mathcal{W}_{\tilde{n}}, \quad (8.35)$$

of  $\tilde{A}\tilde{U} = \tilde{B}$  has a unique solution.

*Proof* Apply the Lax–Milgram theorem (Theorem B.44).  $\blacktriangle$

Theorem 8.1 can be used to bound the discrepancy between weak solutions of  $AU = B$  and  $\tilde{A}\tilde{U} = \tilde{B}$ , that is, the solutions of  $\mathcal{B}(U, W) = E[B'W^*]$ , and  $\tilde{\mathcal{B}}(\tilde{U}, W) = E[\tilde{B}'W^*]$  for all  $W \in \mathcal{W}_{\tilde{n}}$ .

**Theorem 8.9** Let  $\tilde{U}$  be an  $\mathbb{R}^d$ -valued random variable with coordinates  $\tilde{U}_i \in \mathcal{W}_{\tilde{n}}$  given by (8.33). The unknown coefficients in the expression of  $\tilde{U}$  can be calculated from the system of equations  $E[\tilde{A}\tilde{U}\psi_{n_1, n_2, \dots}(G)] = E[\tilde{B}\psi_{n_1, n_2, \dots}(G)]$ , where  $n_k \geq 0$ ,  $n_1 + n_2 + \dots = n$ ,  $n = 0, 1, \dots, \tilde{n}$ , and  $(\tilde{A}, \tilde{B})$  are approximations of  $(A, B)$  obtained by projecting the random elements in these matrices on  $\mathcal{W}_{\tilde{n}}$ .

*Proof* If  $\tilde{U} \in \mathcal{W}_{\tilde{n}}$  is a weak solution for (8.1), it satisfies  $E[(\tilde{A}\tilde{U})'W^*] = E[\tilde{B}'W^*]$  which gives  $\sum_{i,j=1}^d E[\tilde{A}_{ij} \tilde{U}_j W_i^*] = \sum_{i=1}^d E[\tilde{B}_i W_i^*]$  for trial vectors  $W \in \mathcal{W}_{\tilde{n}}$  with



coordinates  $W_i = \sum_{n=0}^{\bar{n}} \sum_{n_1+n_2+\dots=n} d_{n_1,n_2,\dots}^i \psi_{n_1,n_2,\dots}(G)$ . An equivalent form of this condition is

$$\sum_{n=0}^{\bar{n}} \sum_{n_1+n_2+\dots=n} \sum_{i=1}^d d_{n_1,n_2,\dots}^i \left( \sum_{j=1}^d E[\tilde{A}_{ij} \tilde{U}_j \psi_{n_1,n_2,\dots}(G)] - E[\tilde{B}_i \psi_{n_1,n_2,\dots}(G)] \right) = 0,$$

which yields the stated equations for the coefficients in the expression of  $\tilde{U}$  since  $\{d_{n_1,n_2,\dots}^i\}$  are arbitrary. Note that the equations defining the coefficients in the expression of  $\tilde{U}$  constitute projections of  $\tilde{A}\tilde{U} = \tilde{B}$  on  $\psi_{n_1,n_2,\dots}(G)$ , where  $(\tilde{A}, \tilde{B})$  are  $(A, B)$  with  $X$  replaced with  $\tilde{X}$ . ▲

The construction of the polynomial chaos representation  $\tilde{U}$  in (8.33) can pose notable difficulties since the dimension of  $A$  is large in realistic applications,  $\bar{n}$  needs to be relatively large for solution accuracy, and the equations satisfied by the coefficients in the expression of  $\tilde{U}$  are coupled.

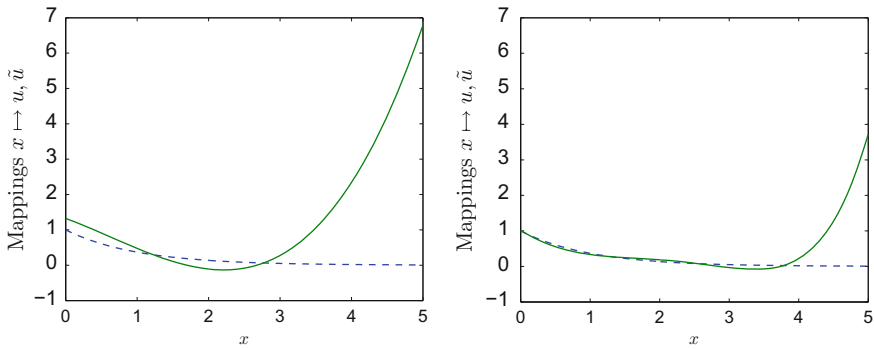
*Example 8.14* Consider (8.1) with  $d = 1$ ,  $A = \exp(X)$ ,  $X \sim N(0, 1)$ , and  $B = 1$ . The exact solution of this equation is  $U = \exp(-X)$ . The approximate solution given by (8.33) has the expression (B.50)

$$\tilde{U} = \sum_{n=0}^{\bar{n}} \frac{1}{\sqrt{n!}} a_n H_n(X), \quad (8.36)$$

depending on the coefficients  $(a_0, a_1, \dots, a_{\bar{n}})$  that can be calculated from a linear system with  $\bar{n} + 1$  equations. We have  $a_0 = 1.1031$ ,  $a_1 = -0.3927$ ,  $a_2 = -0.3126$ , and  $a_3 = 0.2883$  for  $\bar{n} = 3$  and  $a_0 = 1.1031$ ,  $a_1 = 1.5444$ ,  $a_2 = -1.3195$ ,  $a_3 = 0.5655$ ,  $a_4 = -0.2063$  and  $a_5 = 0.0987$  for  $\bar{n} = 5$ . Figure 8.8 shows with dashed and solid lines the exact mapping  $x \mapsto u = \exp(-x)$  and the approximate mappings  $x \mapsto \tilde{u} = \sum_{n=0}^{\bar{n}} a_n H_n(x)/\sqrt{n!}$  given by (8.36) with  $\bar{n} = 3$  (left panel) and  $\bar{n} = 5$  (right panel). The approximate mapping improves with  $\bar{n}$ . Yet, there are notable differences between the exact mapping and its approximations. The mean and standard deviation of  $\tilde{U}$  are 1.1035 and 0.5802 for  $\bar{n} = 3$  and 1.5459 and 1.4539 for  $\bar{n} = 5$ . For  $\bar{n} = 5$ , the mean and standard deviation of  $\tilde{U}$  are in error by  $-6.4\%$  and  $-33.3\%$ , respectively. ◇

*Proof* The lognormal variable  $A$  has the representation  $A = \sum_{m=0}^{\infty} e^{1/2} H_m(X)/m!$  (Example B.44), where  $H_m$  are Hermite polynomials defined by (B.46). The first few Hermite polynomials are  $H_0(x) = 1$ ,  $H_1(x) = x$ ,  $H_2(x) = x^2 - 1$ ,  $H_3(x) = x^3 - 3x$ ,  $H_4(x) = x^4 - 6x^2 + 3$ , and  $H_5(x) = x^5 - 10x^3 + 15x$ . The solution  $U$  can be expressed as  $U = \sum_{n=0}^{\infty} a_n H_n(X)/\sqrt{n!}$ , where the equality holds in the mean square sense. The coefficients  $a_n$  of a truncated version  $\tilde{U} = \sum_{n=0}^{\bar{n}} a_n H_n(X)/\sqrt{n!}$  of  $U$  can be calculated from (Theorem 8.9)

$$\sum_{n=0}^{\bar{n}} \frac{1}{\sqrt{n!}} a_n \sum_{m=0}^{\bar{n}} \frac{1}{m!} E[H_m(X) H_n(X) H_r(X)] = e^{-1/2} E[H_r(X)], \quad r = 0, 1, \dots, \bar{n}, \quad (8.37)$$



**Fig. 8.8** Mappings  $x \mapsto u = \exp(-x)$  (dotted lines) and  $x \mapsto \tilde{u}$  (solid lines) with  $\bar{n} = 3$  (left panel) and  $\bar{n} = 5$  (right panel)

a coupled system of  $\bar{n} + 1$  equations for the unknown coefficients  $(a_0, a_1, \dots, a_{\bar{n}})$ .  $\blacktriangle$

The stochastic Galerkin method is particularly attractive for applications since (1) the approximate solutions  $\tilde{U}$  are represented by finite sums of Hermite polynomials of Gaussian variables with unknown deterministic coefficients, (2) the unknown coefficients in the expression of  $\tilde{U}$  satisfy deterministic equations, and (3) properties of  $\tilde{U}$  can be estimated efficiently from the expression of  $\tilde{U}$  and samples of Gaussian variables since the expression of  $\tilde{U}$  is available.

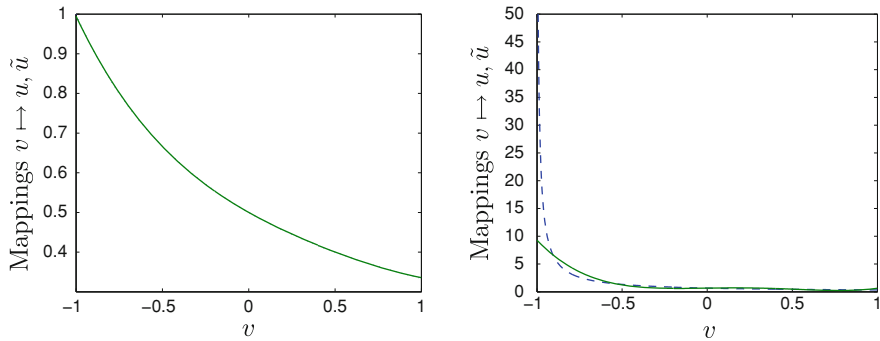
However, the method has notable limitations related primarily to computation demand and accuracy. The computation effort needed to construct a Galerkin solution  $\tilde{U}$  increases rapidly with the order  $\bar{n}$  of the approximation and/or the size of  $X$ . The number of terms in the representation given by (8.33) for each coordinate of  $U$  is equal to  $(d_x + \bar{n})!/(d_x! \bar{n}!)$ , for example, this number is  $(1 + \bar{n})!/\bar{n}! = 1 + \bar{n}$  for  $d_x = 1$  in agreement with (8.36). Also, as previously mentioned, the equations for the unknown coefficients of the polynomial chaos representation for  $\tilde{U}$  are coupled, so that efficient and robust solvers are needed for calculating these coefficients. Since  $d_x$  is usually large in applications, the order  $\bar{n}$  needs to be kept small. This constraint may result in unsatisfactory approximations, as illustrated by Example 8.14. Moreover, even if  $X$  admits a polynomial chaos representation with finite order, the Galerkin solution  $\tilde{U}$  may not belong to the space spanned by the polynomial chaoses representing  $X$ , so that its accuracy may not be satisfactory.

It has also been shown that high expansion order are required when the dependence of the solution  $U$  on the input random parameters  $X$  is not smooth [28]. A related matter is the convergence of the polynomial chaos representations. Under the assumption that  $U = A^{-1}B$  is in  $L^2$ , the sequence of polynomial chaos representations  $\tilde{U}$  converges in m.s. to  $U$  as  $\bar{n} \rightarrow \infty$ , so that it also converges in both probability and distribution. This convergence does not imply that moments of order 3 and higher of  $U$  can be approximated by corresponding moments of  $\tilde{U}$ ,  $\bar{n} < \infty$ , since the sequence of higher order moments of  $\tilde{U}$  may diverge ([9] and Example B.45).

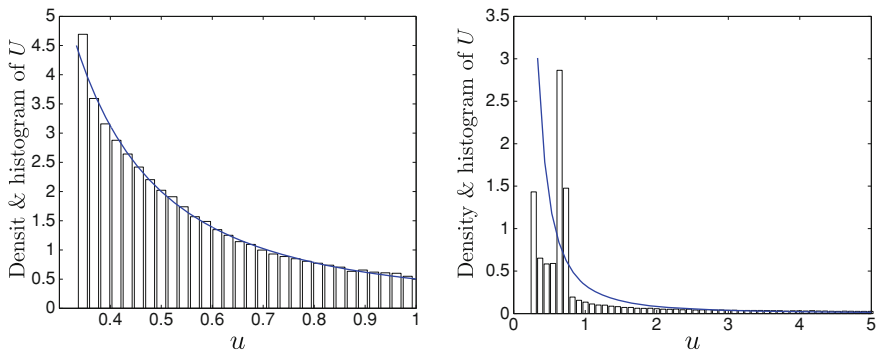
The accuracy of stochastic Galerkin solutions can also be affected by other factors, such as the type of polynomial used to approximate  $X$  and  $U$ . It has been shown that the accuracy of the Galerkin solution can be improved by using Wiener–Askey polynomials of various types depending on the probability law of  $X$ . The improvement relative to Hermite polynomials of Gaussian variables can be significant when dealing with random variables taking values in bounded intervals [29]. This observation can yield efficient and accurate polynomial chaos representations for the input random vector  $X$  but it may be of limited value for the solution  $U$  of (8.1), since the probability law and the range of  $U$  are unknown.

*Example 8.15* Consider the stochastic algebraic equation  $AU = 1$  in Example 8.5 with  $A = X$  and  $X$  uniformly distributed in  $[a, b]$ ,  $0 < a < b < \infty$ . It has been shown that Lagrange, rather than Hermite, polynomials provide an optimal representation for  $X$  [29]. However, representations of solutions using polynomials that are optimal for input parameters may not be optimal for solutions, as illustrated by the following two figures. The solid and dotted lines in Fig. 8.9 are the exact and the approximate mappings from  $v$  to  $u$  and  $v$  to polynomial chaos approximations  $\tilde{u}$  of  $u$  based on Legendre polynomials up to degree 4, where  $V \sim U(-1, 1)$  and  $X = (a+b)/2 + (b-a)V/2$ . Results in the left and the right panels are for  $(a, b) = (1, 3)$  and  $(a, b) = (0.01, 3)$ , respectively. The mappings in the left panel are indistinguishable at the scale of the figure, but differ significantly in the right panel. The density of  $U$  and histograms of  $\tilde{U}$  are shown in Fig. 8.10 for  $(a, b) = (1, 3)$  (left panel) and  $(a, b) = (0.01, 3)$  (right panel). The histogram of  $\tilde{U}$  matches accurately and differs significantly from the density of  $U$  for  $(a, b) = (1, 3)$  and  $(a, b) = (0.01, 3)$ , respectively. The mean and the standard deviation of  $\tilde{U}$  nearly coincide with corresponding moments of  $U$  for  $(a, b) = (1, 3)$  but for  $(a, b) = (0.01, 3)$  are in error by  $-23.19\%$  and  $-64.35\%$ . These numerical results are consistent with the quality of the mapping  $v \mapsto \tilde{u}$  in Fig. 8.9, and relate to differences between the support and the shape of the distributions of  $X$  and  $U$ . For  $(a, b) = (0.01, 3)$ , the mapping  $v \mapsto \tilde{u}$  is inaccurate, for example, the range of  $U$  is  $[1/b, 1/a] = [1/3, 100]$  while, the range on  $\tilde{U}$  is  $[0.2374, 9.2797]$  for  $\tilde{n} = 4$ , so that  $U$  and  $\tilde{U}$  cannot have similar distributions. Also, Legendre polynomial representations are not optimal for  $U$  since its distribution differs significantly from a uniform distribution, so that approximate solutions based on these polynomials are not likely to be satisfactory.  $\diamond$

*Proof* Legendre polynomials are defined by  $\phi_i(v) = d^i((v+1)^i(v-1)^i)/dv^i$ ,  $i = 0, 1, \dots$ ,  $v \in (-1, 1)$ , are orthogonal, that is,  $\int_{-1}^1 \phi_i(v)\phi_j(v)dv = 0$  for  $i \neq j$ , and have norm  $\|\phi_i\|^2 = \int_{-1}^1 \phi_i(v)^2 dv/2 = (i!)^2 2^{2i}/(2i+1)$  ([22], Chap. 12). Note that  $X = (a+b)/2 + (b-a)V/2 = [(a+b)/2]\phi_0(V) + [(b-a)/4]\phi_1(V)$ , where  $v$  is uniformly distributed in  $(-1, 1)$ . Consider the approximate representation  $\tilde{U} = \sum_{i=0}^{\tilde{n}} \beta_i \phi_i(V)$  of  $U$ . The unknown coefficients  $\{\beta_i\}$  can be obtained by projecting the approximate version  $\tilde{A}\tilde{U} = 1$  of  $AU = 1$  on  $\phi_k$ ,  $k = 0, 1, \dots, \tilde{n}$ , which gives the system of linear equations



**Fig. 8.9** Mappings  $v \mapsto u$  (solid lines) and  $v \mapsto \tilde{u}$  (dotted lines) for  $A = X$  uniformly distributed in  $(a, b)$ , where  $(a, b) = (1, 3)$  (left panel), and  $(a, b) = (0.01, 3)$  (right panel)



**Fig. 8.10** Density of  $U$  and histograms of  $\tilde{U}$  for  $(a, b) = (1, 3)$  (left panel) and  $(a, b) = (0.01, 3)$  (right panel)

$$\sum_{i=0}^{\bar{n}} \left( \frac{a+b}{2} E[\phi_i(V)\phi_k(V)] + \frac{b-a}{2} E[V\phi_i(V)\phi_k(V)] \right) \beta_i = E[\phi_k(V)], \quad k = 0, 1, \dots, \bar{n},$$

for  $(\beta_0, \dots, \beta_p)$ . Resulting values of these coefficients and the representation  $\tilde{U} = \sum_{i=0}^{\bar{n}} \beta_i \phi_i(V)$  can be used in conjunction with Monte Carlo simulation to find statistics of  $U$  approximately. ▲

### 8.2.5 Stochastic Collocation Method

As for Monte Carlo simulation and SROMs, the implementation of the stochastic collocation method involves solutions of deterministic versions of (8.1) for specified values of  $X$ . These solutions exist if  $A$  has an inverse a.s., so that we assume that  $A^{-1}$  exists with probability 1. Let  $(x^{(0)}, x^{(1)}, \dots, x^{(n)})$  be  $(n+1)$  distinct points in the range  $\Gamma = X(\Omega)$  of the random vector  $X$  collecting the random parameters in the

definition of matrices  $A$  and  $B$  in (8.1), referred to as collocation points. Consider the approximation

$$\tilde{U} = \sum_{i=0}^n u^{(i)} \ell_i(X) \quad (8.38)$$

for  $U$ , where  $u^{(i)}$  are solutions of (8.1) for  $X$  set equal to  $x^{(i)}$  and  $\ell_i(X)$  are interpolating polynomials centered at  $x^{(i)}$ ,  $i = 0, 1, \dots, n$ . Note that the accuracy of  $\tilde{U}$  depends on the number and the location of the collocation points  $(x^{(0)}, x^{(1)}, \dots, x^{(n)})$ , the type of interpolating polynomials, the properties of mapping  $X \mapsto U$ , and the probability law of  $X$ . Properties of  $\tilde{U}$  can be obtained simply and efficiently from its expression by, for example, Monte Carlo simulation since the mapping  $X \mapsto \tilde{U}$  is available analytically. The collocation method can be viewed as a response surface that approximates the mapping  $X \mapsto U$  from deterministic solutions of (8.1). Note that  $\tilde{U}$  in (8.38) and  $\tilde{U}$  in previous sections have different meaning.

If the mapping  $X \mapsto U$  defined by (8.1) is continuous, then it is possible to construct a polynomial approximation  $\tilde{U}$  for  $U$  that has a specified accuracy, as indicated by the following theorem stated for  $d = 1$ .

**Theorem 8.10** *For  $g \in C[a, b]$  and  $\varepsilon > 0$ , there exists an algebraic polynomial  $p$  such that  $\|g - p\|_\infty = \max_{a \leq x \leq b} |g(x) - p(x)| \leq \varepsilon$  ([22], Theorem 6.1, the Weierstrass theorem).*

*Example 8.16* Consider the stochastic algebraic equation in Example 8.5 and let  $u^{(i)} = 1/x^{(i)}$  be the solution of  $AU = 1$  for  $A = X$  set equal to  $x^{(i)} = a + i(b-a)/n$ ,  $i = 0, 1, \dots, n$ . The approximation  $\tilde{U}$  of  $U$  given by (8.38) is

$$\tilde{U} = \sum_{i=0}^n \frac{1}{a + i(b-a)/n} \prod_{j=0, j \neq i}^n \frac{X - x^{(j)}}{x^{(i)} - x^{(j)}} = \sum_{i=0}^n \frac{1}{a + i(b-a)/n} \ell_i(X), \quad (8.39)$$

where  $\ell_i(X) = \prod_{j=0, j \neq i}^n (X - x^{(j)}) / (x^{(i)} - x^{(j)})$  are Lagrange interpolating polynomials. The approximate solution  $\tilde{U}$  matches  $U$  exactly at the collocation points since  $\ell_i(x^{(j)}) = \delta_{ij}$ . Moreover, for every  $\varepsilon > 0$ , there is a polynomial approximation  $\tilde{U}$  such that  $E[|U - \tilde{U}|^q] \leq \varepsilon^q$ , where  $q \geq 1$  is an integer.  $\diamond$

*Proof* The mapping  $x \mapsto u(x) = 1/x$  is continuous for  $x \neq 0$ , so that there is a polynomial  $\tilde{u}(x)$  such that  $\max_{a \leq x \leq b} |u(x) - \tilde{u}(x)| \leq \varepsilon$  (Theorem 8.10). Accordingly,  $E[|U - \tilde{U}|^q] \leq E[\varepsilon^q] = \varepsilon^q$ .  $\blacktriangle$

**Theorem 8.11** *Let  $g : [a, b] \rightarrow \mathbb{R}$  and  $p$  be as in Theorem 8.10,  $X$  a real-valued random variable with support  $[a, b]$ , and  $q \geq 0$  an integer. There exists a constant  $c > 0$  such that*

$$|E[g(X)^{q+1}] - E[p(X)^{q+1}]| \leq c \|g - p\|_\infty. \quad (8.40)$$

*Proof* The identity  $\alpha^{q+1} - \beta^{q+1} = (\alpha - \beta) \sum_{k=0}^q \alpha^{q-k} \beta^k$  applied to  $E[g(X)^{q+1}] - E[p(X)^{q+1}] = E[g(X)^{q+1} - p(X)^{q+1}]$  gives

$$\begin{aligned}
|E[g(X)^{q+1} - p(X)^{q+1}]| &\leq E\left[|g(X) - p(X)| \sum_{k=0}^q g(X)^{q-k} p(X)^k\right] \\
&\leq \|g - p\|_\infty E\left[\sum_{k=0}^q g(X)^{q-k} p(X)^k\right] = c\|g - p\|_\infty,
\end{aligned}$$

where  $c = E\left[\sum_{k=0}^q g(X)^{q-k} p(X)^k\right] > 0$  is a finite constant since  $X$  has bounded support and  $g$  and  $p$  are continuous in  $[a, b]$  and, therefore, bounded in this interval. The bound in (8.40) shows that moments of  $g(X)$  can be approximated by corresponding moments of  $p(X)$  if  $\|g - p\|_\infty$  is sufficiently small.  $\blacktriangle$

**Theorem 8.12** *Let  $g : [a, b] \rightarrow \mathbb{R}$  be a continuous function and  $p$  a polynomial of degree  $n$  satisfying the interpolation conditions  $p(x^{(i)}) = g(x^{(i)})$ , where  $\{x^{(i)}\}$ ,  $i = 0, 1, \dots, n$ , are distinct points in  $[a, b]$ . If  $g \in C^{(n+1)}[a, b]$ , the error  $e_n(x) = g(x) - p(x)$  is*

$$e_n(x) = \frac{1}{(n+1)!} \prod_{i=0}^n (x - x^{(i)}) g^{(n+1)}(\xi), \quad \xi \in [a, b], \quad (8.41)$$

and  $\lim_{n \rightarrow \infty} e_n(x) = 0$  for all  $x \in [a, b]$ .

*Proof* For proof, see [22] (Sect.4.2). We only show the convergence  $e_n(x) \rightarrow 0$  as  $n \rightarrow \infty$ . Since  $g^{(n+1)}$  is continuous by assumption,  $M = \max_{a \leq x \leq b} |g^{(n+1)}(x)|$  is finite, so that  $|e_n(x)| \leq (b-a)^{n+1} M / (n+1)!$  by (8.41). Let  $n^*$  be such that  $n^* < b-a \leq n^* + 1$  so that, for  $n > n^*$ ,

$$|e_n(x)| \leq \frac{(b-a)^{n^*} M}{(n^*)!} \prod_{k=n^*+1}^n \frac{b-a}{k+1},$$

so that  $|e_n(x)| \rightarrow 0$  as  $n \rightarrow \infty$ .  $\blacktriangle$

*Example 8.17* Let  $p(X) = \sum_{i=1}^n (1/x^{(i)}) \ell_i(X)$  be a polynomial approximation for random variable  $g(X) = 1/X$ , where  $x^{(i)}$  and  $\ell_i$  are as in (8.39) and  $X$  is uniformly distributed in  $[a, b]$ ,  $0 < a < b < \infty$ . The discrepancy between moments of  $g(X)$  and  $p(X)$  decreases with the degree  $n$  of  $p(X)$  in agreement with Theorem 8.11. For example, if  $a = 0.1$  and  $b = 1$ , the errors of  $E[p(X)^6]$  relative to  $E[g(X)^6]$  is 29.18%, 2.84%, 0.11%, and 0.07% for  $n = 5, 10, 20$ , and 50, respectively. These errors are based on estimates of  $E[p(X)^6]$  obtained from  $10^6$  independent samples of  $X$ .  $\diamond$

A particularly useful class of interpolation polynomials are the Bernstein polynomials, that have been originally introduced to provide an alternative proof of the Weierstrass theorem ([19], Chap. 1).

**Definition 8.3** For a real-valued function  $g$  defined on the closed interval  $[0, 1]$ ,

$$B_n(x) = \sum_{i=0}^n g(i/n) \frac{n!}{i!(n-i)!} x^i (1-x)^{n-i} = \sum_{i=0}^n g(i/n) p_{n,i}(x), \quad x \in [0, 1], \quad (8.42)$$

is called the Bernstein polynomial of order  $n$  for function  $g$ .

The polynomials  $p_{n,i}(x) = [n!/(i!(n-i)!)]x^i(1-x)^{n-i}$  in the expression of  $B_n(x)$  are positive and satisfy the condition  $\sum_{i=0}^n p_{n,i}(x) = 1$  for all  $x \in [0, 1]$ . The latter property implies  $m \leq B_n(x) \leq M$  if  $m \leq g(x) \leq M$ ,  $\forall x \in [0, 1]$ .

**Theorem 8.13** *If  $g : [0, 1] \rightarrow \mathbb{R}$  is a bounded function, then  $\lim_{n \rightarrow \infty} B_n(x) = g(x)$  at each point of continuity  $x$  of  $g$ . If  $g$  is continuous in  $[0, 1]$ , then  $\lim_{n \rightarrow \infty} B_n(x) = g(x)$  holds uniformly in  $[0, 1]$  ([19], Theorem 1.1.1).*

The Weierstrass theorem is a direct consequence of Theorem 8.13. That  $g$  has support  $[0, 1]$  is not restrictive since any interval  $[a, b]$  can be mapped into  $[0, 1]$  by a linear mapping. For example,

$$B_n(x; b) = \sum_{i=0}^n g(bi/n) p_{n,i}(x/b), \quad x \in [0, b], \quad (8.43)$$

is a Bernstein polynomial for a bounded function  $g : [0, b] \rightarrow \mathbb{R}$ . The following two theorems are extensions of the Bernstein polynomial representation in Theorem 8.13 to functions defined on unbounded intervals and multivariate functions.

**Theorem 8.14** *Let  $g : [0, \infty) \rightarrow \mathbb{R}$  be a bounded function and  $b_n \sim o(n)$ . Then  $\lim_{n \rightarrow \infty} B_n(x; b_n) = g(x)$  at any point of continuity of  $g$  ([19], Theorem 2.3.1).*

**Theorem 8.15** *Let  $g : [0, 1]^{d_x} \rightarrow \mathbb{R}$  be a bounded function. Then the multivariate Bernstein polynomial*

$$\begin{aligned} & B_{n_1, \dots, n_{d_x}}(x_1, \dots, x_{d_x}) \\ &= \sum_{i_1=0}^{n_1} \cdots \sum_{i_{d_x}=0}^{n_{d_x}} g\left(\frac{i_1}{n_1}, \dots, \frac{i_{d_x}}{n_{d_x}}\right) \prod_{j=1}^{d_x} \frac{n_j!}{i_j!(n_j - i_j)!} p_{n_j, i_j}(x_j), \quad x \in \mathbb{R}^{d_x}, \end{aligned} \quad (8.44)$$

*has the property  $\lim_{n_1, \dots, n_{d_x} \rightarrow \infty} B_{n_1, \dots, n_{d_x}}(x_1, \dots, x_{d_x}) = g(x_1, \dots, x_{d_x})$  at all points of continuity of  $g$ . If  $g$  is continuous in  $[0, 1]^{d_x}$ , then  $\lim_{n_1, \dots, n_{d_x} \rightarrow \infty} B_{n_1, \dots, n_{d_x}}(x_1, \dots, x_{d_x}) = g(x_1, \dots, x_{d_x})$  holds uniformly in  $[0, 1]^{d_x}$  ([14], [19], Sect. 2.9).*

This result shows that there is an analogue of the Weierstrass theorem (Theorem 8.10) in the multidimensional case. Theorem 8.15 states that for any real-valued continuous function  $g$  defined on a bounded rectangle  $R$  in  $\mathbb{R}^{d_x}$  and for any  $\varepsilon > 0$ , there exists an algebraic polynomial  $p$  such that  $\|g - p\|_\infty = \max_{x \in R} |g(x) - p(x)| \leq \varepsilon$ . Moreover, Theorem 8.11 holds in the multivariate case, that is, if  $X$  is an  $\mathbb{R}^{d_x}$ -valued random variable with support  $R$ , the moments of any order of  $g(X)$  can be approximated by moments of corresponding order of  $p(X)$  at any desired accuracy.

The latter observations imply that, if  $U$  in (8.1) is a continuous function of  $X$  and  $X$  has a bounded range  $\Gamma = X(\Omega)$ , then  $U$  can be approximated to any degree of accuracy by a polynomial  $\tilde{U}$  of  $X$  with coefficients obtained from solutions of (8.1) with  $X$  set equal to collocation points selected in  $\Gamma = X(\Omega)$ , that is, the condition  $\|U - \tilde{U}\|_\infty \leq \varepsilon$  can be satisfied for arbitrary  $\varepsilon > 0$ . Accordingly,  $\tilde{U}$  converges a.s. to  $U$  as its degree increases indefinitely, and moments of  $U$  can be approximated by moments of  $\tilde{U}$  for a sufficiently large polynomial degree. The rate of convergence of  $\tilde{U}$  to  $U$  depends on properties of  $X$  and of the mapping  $X \mapsto U$  as well as the number and location of collocation points.

It is not possible to find the degree of an approximating polynomial for a required accuracy since the mapping  $X \mapsto U$  is not known. Note that the construction of  $\tilde{U}$  involves  $\prod_{j=1}^{d_x} n_j$  solutions of distinct deterministic versions of (8.1), where  $n_j$  denotes the number of collocation points along coordinate  $j = 1, \dots, d_x$ , so that the degree of  $\tilde{U}$  needs to be relatively low unless the dimension  $d_x$  of  $X$  is small. For example,  $\prod_{j=1}^{d_x} n_j = 10^{100}$  for  $n_j = 10$  and  $d_x = 100$ , so that this version of the method may be impractical for many applications. Alternative collocation schemes based on the Smolyak formula and other schemes have been proposed to reduce calculations related to the solution of stochastic elliptic partial differential equations [21]. Additional information on these schemes, a priori bounds on the discrepancy between exact and collocation solutions of (8.1), and comparisons between stochastic Galerkin and collocation methods can be found elsewhere [2, 3].

*Example 8.18* Consider the scalar stochastic algebraic equation  $AU = 1$  in Example 8.16 with  $A = X$  uniformly distributed in  $(a, b)$ . The approximate mapping  $X \mapsto \tilde{U}$  is given by (8.39). The errors in percentages between the exact and approximate means are 0.085%, 0.026%, and 0.0259% for  $(a, b) = (1, 3)$  and  $n = 4, 10$ , and 20, respectively, and 347.12%, 97.12%, and 30.84% for  $(a, b) = (0.01, 3)$  and the same values of  $n$ . The corresponding errors of approximate standard deviations are 0.703%, 0.0129%, and 0.0123% for  $(a, b) = (1, 3)$  and 290.99%, 128.39%, and 58.12% for  $(a, b) = (0.01, 3)$ . Improvements result by increasing the degree of the interpolating polynomials. For example, the mean and standard deviation of  $\tilde{U}$  are in error by 7.39% and 19.44%, respectively, for  $(a, b) = (0.01, 3)$  and  $n = 40$ .

The solution of  $AU = 1$  with  $A \sim U(0.01, 3)$  in Example 8.5 using a SROM with  $m = 20$  provides superior approximations for the mean and standard deviation of  $U$ . The errors of these approximations are 2.76% and  $-15.89\%$ , respectively.  $\diamond$

### 8.2.6 Reliability Method

We have seen that (8.30) can be used to map an arbitrary non-Gaussian  $\mathbb{R}^{d_x}$ -valued random variable  $X$  into a Gaussian vector  $G = (G_1, \dots, G_{d_x})$  with independent  $N(0, 1)$  coordinates, so that  $X$  in the definition of (8.1) can be assumed to coincide with  $G$  without loss of generality.



A common objective in reliability studies is the calculation of the probability

$$p_s = P(h(U) \leq 0) = P(h(U(G)) \leq 0) = P(G \in D) = \int_D \phi(x) dx, \quad (8.45)$$

where  $D = \{x \in \mathbb{R}^{d_x} : g(x) \leq 0\}$  denotes a safe set,  $g(x) = h(u(x))$ ,  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  is a measurable function, and  $\phi(x) = \prod_{j=1}^{d_x} \exp(-x_j^2/2) \sqrt{2\pi}$ .

Generally, the integral  $\int_D \phi(x) dx$  cannot be obtained analytically. Numerical solutions, for example, Monte Carlo simulation or quadratures, are usually inefficient when  $d_x$  is large. Alternative solutions are offered by reliability methods. The simplest reliability method approximates  $p_s$  by

$$p_s \simeq \Phi(\beta), \quad (8.46)$$

where  $\Phi$  denotes the distribution of  $N(0, 1)$  and  $\beta$  is the minimum distance from the origin of  $\mathbb{R}^{d_x}$  to the boundary of  $D$ . Theoretical considerations regarding the construction of (8.46) and related approximations can be found in [8] (Chap. 9). Note that the approximation in (8.46) replaces the calculation of the multidimensional integral  $\int_D \phi(x) dx$  in (8.45) with the solution of a constraint optimization problem that finds the point  $x^* \in \mathbb{R}^{d_x}$  on the boundary  $\partial D = \{x \in \mathbb{R}^{d_x} : g(x) = 0\}$  of  $D$  that has the smallest norm. Details on optimization algorithms for finding  $x^*$  and numerical examples can be found in [26].

The reliability method differs from the other methods for solving stochastic algebraic equations in both focus and approach. The method has been developed to calculate probabilities that the solution  $U$  of (8.1) belongs to a specified subset of  $\mathbb{R}^d$ . Calculation of moments of  $U$  by this method is usually inefficient. The method is based on asymptotic properties of multidimensional integrals of Gaussian probability density functions [5]. Useful numerical comparisons between the accuracy of the stochastic Galerkin and the reliability methods can be found in [26].

### 8.3 SAEs with Small Uncertainty

Throughout this section we consider the special case of (8.1) in which  $B$  is deterministic and  $A$  depends on an  $\mathbb{R}^{d_x}$ -valued random variable  $X$  that has small uncertainty. The assumption that  $B$  is deterministic is not restrictive. We apply the Taylor series, perturbation series, Neumann series, and equivalent linearization methods to solve approximately this class of stochastic algebraic equations. The focus is on the first two moments of the solution  $U$  of (8.1).

### 8.3.1 Taylor Series

Let  $\mu_x = \{\mu_{x,p}\}$  and  $\gamma_x = \{\gamma_{x,pq}\}$ ,  $p, q = 1, \dots, d_x$ , denote the mean and the covariance matrices of  $X$ . If the mapping  $X \mapsto U(X)$  has continuous second order partial derivatives, then

$$U(X) \simeq \tilde{U}(X) = U(\mu_x) + \sum_{p=1}^{d_x} \frac{\partial U(\mu_x)}{\partial x_p} (X_p - \mu_{x,p}) \quad (8.47)$$

by using the first two terms of the Taylor expansion of  $U(X)$  about the mean of  $X$ , where  $\partial U(\mu_x)/\partial x_p$  is a vector with coordinates  $\partial U_i(\mu_x)/\partial x_p$ ,  $i = 1, \dots, d$ .

**Theorem 8.16** *The approximate mean and covariance matrices of  $U$  given by its representation in (8.47) are*

$$\begin{aligned} \mu_u &\simeq U(\mu_x) \\ \gamma_u &\simeq \sum_{p,q=1}^{d_x} \frac{\partial U(\mu_x)}{\partial x_p} \left( \frac{\partial U(\mu_x)}{\partial x_q} \right)' \gamma_{x,pq}, \end{aligned} \quad (8.48)$$

and depend on the first two moments of  $X$ .

*Proof* The Taylor expansion of  $U(X)$  about  $x^{(0)} \in \mathbb{R}^{d_x}$  is

$$U(X) = U(x^{(0)}) + \sum_{p=1}^{d_x} \frac{\partial U(x^{(0)})}{\partial x_p} (X_p - x_p^{(0)}) + R_2(X, x^{(0)}),$$

where

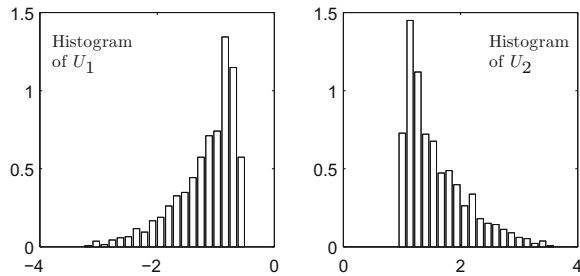
$$R_2(X, x^{(0)}) = \frac{1}{2} \sum_{p,q=1}^{d_x} \frac{\partial^2 U(X^*)}{\partial x_p \partial x_q} (X_p - x_p^{(0)}) (X_q - x_q^{(0)})$$

and  $X^* = \theta x^{(0)} + (1 - \theta)X$ ,  $\theta \in (0, 1)$  ([4], Theorem 21.1). The approximation in (8.47) is the above Taylor expansion for  $x^{(0)} = \mu_x$  without the remainder  $R_2$ . The error  $R_2(X, \mu_x) = U(X) - \tilde{U}(X)$  of  $\tilde{U}(X)$  in (8.47) depends on properties of  $X$  and of mapping  $X \mapsto U(X)$  in a neighborhood of  $\mu_x$ . ▲

**Theorem 8.17** *The second moment properties of  $\tilde{U}(X)$  in (8.48) can be calculated from*

$$U(\mu_x) = A(\mu_x)^{-1} B \quad \text{and} \quad (8.49)$$

$$\frac{\partial U(\mu_x)}{\partial x_u} = A(\mu_x)^{-1} \left( - \frac{\partial A(\mu_z)}{\partial x_u} U(\mu_x) \right). \quad (8.50)$$

**Fig. 8.11** Histograms of  $U_1$  and  $U_2$ 

*Proof* The solution of (8.1) with  $X = \mu_x$  gives  $U(\mu_x)$ . The derivative,

$$\frac{\partial A(X)}{\partial X_p} U(X) + A(X) \frac{\partial U(X)}{\partial X_p} = \frac{\partial B}{\partial X_p},$$

of (8.1) with respect to coordinate  $X_p$  of  $X$  gives (8.50) for  $X = \mu_x$  since  $\partial B / \partial X_p = 0$  by assumption. The rates of change  $\partial U(\mu_x) / \partial x_u$  of  $U$  relative to the coordinates of  $X$  are referred to as sensitivity factors. If the sensitivity factor of  $U$  with respect to a coordinate  $X_p$  of  $X$  and the uncertainty in  $X_p$  is small relative to the other coordinates of  $X$ , the uncertainty in  $X_p$  has a minor contribution to the uncertainty in  $U$  so that we can set  $X_p = \mu_{x,p}$ .  $\blacktriangle$

Previous observations show that approximation of  $U$  in (8.47) has the form

$$U \simeq \tilde{U}(X) = A(\mu_x)^{-1} \left[ I - \sum_{p=1}^{d_x} \frac{\partial A(\mu_x)}{\partial x_p} A(\mu_x)^{-1} (X_p - \mu_{x,p}) \right] B, \quad (8.51)$$

where  $I$  denotes the  $(d, d)$  identity matrix. The extension to the case in which  $A$  and  $B$  depend on  $X$  results by similar arguments (Exercise 8.11).

*Example 8.19* Let  $U = (U_1, U_2)$  be the solution of (8.1) with

$$A = \begin{bmatrix} \cos(\Theta_1) & \cos(\Theta_2) \\ \sin(\Theta_1) & \sin(\Theta_2) \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 \\ q \end{bmatrix}, \quad (8.52)$$

where  $\Theta_1 = \tan^{-1}(X_2/(X_1 + a))$ ,  $\Theta_2 = \tan^{-1}((a - X_2)/(X_1 + a))$ ,  $X_1$  and  $X_2$  are independent random variables uniformly distributed in  $(-\varepsilon a, \varepsilon a)$ , and  $0 \leq \varepsilon \leq 1/3$ . This stochastic algebraic equation is the equilibrium condition for a system with random imperfections whose solution  $U = (U_1, U_2)$  gives forces in the system. If  $\varepsilon = 0$ , that is, the system has no imperfections,  $U_1 = -1$  and  $U_2 = \sqrt{2}$ .

Figure 8.11 shows histograms of  $U_1$  and  $U_2$  obtained from 1000 independent realizations of the system for  $a = 1$ ,  $\varepsilon = 0.3$ , and  $q = 1$ . The estimated mean, standard deviation, skewness, and kurtosis are  $-1.16$ ,  $0.52$ ,  $1.23$ , and  $4.17$  for  $U_1$  and  $1.58$ ,  $0.52$ ,  $1.18$ , and  $3.94$  for  $U_2$ .

The approximate means of  $U_1$  and  $U_2$  given by the first order Taylor representation in (8.49) are  $-1$  ( $-14\%$ ) and  $\sqrt{2}$  ( $-10.65\%$ ). The corresponding standard

deviations derived from (8.48) and (8.50) are 0.6 (14.81%) and 0.49 (−6.42%). The numbers in parentheses give errors relative to the Monte Carlo simulation.  $\diamond$

The approximate moments of  $U$  in (8.48) can be improved by retaining additional terms from the Taylor series representation of the mapping  $X \mapsto U(X)$ . Under the assumption that  $X \mapsto U(X)$  has continuous third order partial derivatives,  $U$  can be approximated by

$$\begin{aligned} U(X) \simeq \tilde{U}(X) = & U(\mu_x) + \sum_{u=1}^{d_x} \frac{\partial U(\mu_x)}{\partial x_u} (X_u - \mu_{x,u}) \\ & + \frac{1}{2} \sum_{u,v=1}^{d_x} \frac{\partial^2 U(\mu_x)}{\partial x_u \partial x_v} (X_u - \mu_{x,u})(X_v - \mu_{x,v}). \end{aligned} \quad (8.53)$$

The calculation of the second moment properties of  $U$  based on this approximation requires information on  $X$  beyond its first two moments (Exercise 8.13).

Nonlinear stochastic algebraic equations can also be solved approximately by the Taylor series method. Let  $U$  be the solution of

$$AU + N = B, \quad (8.54)$$

where  $A$  is as in (8.1) and  $N$  is a  $d$ -dimensional vector whose coordinates are nonlinear real-valued functions of  $(X, U)$ . If (8.47) is used to calculate the second moment properties of  $U$ , we will need  $U(\mu_x)$  and  $\{\partial U(\mu_x)/\partial x_p\}$ . Note that  $U(\mu_x)$  is the solution of (8.54) with  $X = \mu_x$  and that  $\partial U(\mu_x)/\partial x_p$  can be obtained from the derivatives of (8.54) with respect to the coordinates  $X_p$  of  $X$ , that is,

$$\frac{\partial A}{\partial X_p} U + A \frac{\partial U}{\partial X_p} + \frac{\partial N}{\partial X_p} + \sum_{i=1}^d \frac{\partial N}{\partial U_i} \frac{\partial U_i}{\partial X_p} = 0,$$

for  $X = \mu_x$ . The partial derivatives of  $U$  with respect to  $X_p$  can be calculated since the functions  $X \mapsto A(X)$  and  $(X, U) \mapsto N(X, U)$  are known.

### 8.3.2 Perturbation Series

Let  $A = \bar{A} + \varepsilon R$  in (8.1), where  $\bar{A} = E[A]$  and  $R$  are  $(d, d)$  deterministic and random matrices and  $\varepsilon$  is a small parameter. As previously, it is assumed that  $X$  and  $R$  have finite first two moments. Note that  $E[R_{ij}] = 0$ ,  $i, j = 1, \dots, d$ , by construction.

**Theorem 8.18** *The perturbation solution of (8.1) is*

$$\begin{aligned} U &= U^{(0)} + \varepsilon U^{(1)} + \varepsilon^2 U^{(2)} + O(\varepsilon^3), \quad \text{where} \\ U^{(r)} &= -\bar{A}^{-1} R U^{(r-1)}, \quad r = 1, 2, \dots, \text{ and } U^{(0)} = \bar{A}^{-1} B, \end{aligned} \quad (8.55)$$

so that the approximate first two moments of  $U$  are

$$\begin{aligned} E[U] &= \bar{A}^{-1}B + \varepsilon^2 \bar{A}^{-1}E[R\bar{A}^{-1}R]\bar{A}^{-1}B + O(\varepsilon^3), \\ E[UU'] &= E[U^{(0)}(U^{(0)})'] + \varepsilon^2 E[U^{(0)}(U^{(2)})' + U^{(1)}(U^{(1)})' \\ &\quad + U^{(2)}(U^{(0)})'] + O(\varepsilon^3). \end{aligned} \quad (8.56)$$

*Proof* Since the random part of  $A$  is small, it is expected that the solution  $U$  does not differ significantly from the solution  $\bar{A}^{-1}B$  of (8.1) with  $\varepsilon = 0$ . If the perturbation solution in (8.55) is not singular ([15], Sect. 1.2), the representation of  $U$  in (8.55) holds, and (8.1) takes the form

$$\begin{aligned} (\bar{A} + \varepsilon R)(U^{(0)} + \varepsilon U^{(1)} + \varepsilon^2 U^{(2)} + \dots) &= B \quad \text{or} \\ (\bar{A}U^{(0)} - B) + \varepsilon(\bar{A}U^{(1)} + RU^{(0)}) + \varepsilon^2(\bar{A}U^{(2)} + RU^{(1)}) + \dots &= 0. \end{aligned}$$

This power series in  $\varepsilon$  is zero if  $\bar{A}U^{(0)} = B$  and  $\bar{A}U^{(r)} = -RU^{(r-1)}$  for  $r = 1, 2, \dots$  by a fundamental theorem of perturbation theory ([24], p. 12). Hence, the perturbation solution of order  $p \geq 1$  has the expression

$$\tilde{U}^{(p)} = \sum_{k=0}^p (-1)^k \varepsilon^k (\bar{A}^{-1}R)^k \bar{A}^{-1}B, \quad (8.57)$$

where  $(\bar{A}^{-1}R)^k$  is the identity matrix for  $k = 0$ . Note that the equations for  $U^{(r)}$ ,  $r = 0, 1, \dots$ , have the same deterministic operator. ▲

*Example 8.20* Let  $A$  and  $B$  in (8.52) be matrices defining the stochastic algebraic equation in (8.1), and let  $\bar{A} = E[A]$  and  $\varepsilon R = A - \bar{A}$  denote the deterministic and random parts of  $A$ , respectively. The approximate means of  $U_1$  and  $U_2$  based on the first order perturbation  $\tilde{U}^{(1)}$  are  $-1.0130$  (1.30%) and  $1.4245$  (0.73%) for  $\varepsilon = 0.3$ . The corresponding standard deviations of  $U_1$  and  $U_2$  are  $0.3825$  (−26.81%) and  $0.3831$  (−26.82%). The numbers in parentheses give errors relative to the Monte Carlo solution in Example 8.19. ◇

The calculation of moments of  $U$  based on its representation in (8.57) involves expectations of powers of  $\bar{A}^{-1}R$  that can be calculated efficiently by Monte Carlo simulation. Note also that expansions  $\tilde{U}^{(p)}$  of different orders need to be used to calculate moments of  $U$  with the same accuracy. For example, the error of the mean of  $\tilde{U}^{(1)}$  is of order  $\varepsilon^2$ . To obtain an approximation of the same order for the standard deviation of a coordinate of  $U$ , we need to use the perturbation solution  $\tilde{U}^{(2)}$ . The perturbation method can be applied under some conditions to nonlinear equations of the type in (8.54) but its use is rather convoluted. Note also that the assumption that  $B$  is deterministic is not restrictive since the approximation  $\tilde{U}^{(p)}$  of  $U$  in (8.57) holds for  $B$  random.

### 8.3.3 Neumann Series

The Taylor series and the perturbation methods calculate moments of the solution  $U$  of (8.1) approximately without finding  $A^{-1}$ . The Neumann series method constructs an approximation for  $A^{-1}$ , and uses this approximation to calculate properties of  $U$ . The construction of an approximation for  $A^{-1}$  is based on the following result.

**Theorem 8.19** *If  $C$  is an  $(d, d)$ -deterministic matrix such that  $\|Cx\| \leq \gamma\|x\|$ ,  $0 < \gamma < 1$ , for all  $x \in \mathbb{R}^d$ , then the series*

$$(I + C)^{-1} = \sum_{r=0}^{\infty} (-1)^r C^r \quad (8.58)$$

*is absolutely convergent, where  $I$  denotes the  $(d, d)$  identity matrix,  $C^0 = I$ ,  $\|x\|$  is the Euclidean norm of  $x \in \mathbb{R}^d$  ([27], Chap. 2).*

*Proof* The sequence  $S_k x = \sum_{r=0}^k (-1)^r C^r x$  is Cauchy in  $\mathbb{R}^d$  since, for  $k > l$ ,

$$\|S_k x - S_l x\| \leq \sum_{r=l+1}^k \|C^r x\| \leq \sum_{r=l+1}^k \gamma^r \|x\| = \frac{\gamma^{l+1} - \gamma^{k+1}}{1 - \gamma} \|x\| \rightarrow 0$$

as  $k, l \rightarrow \infty$ . Since  $\mathbb{R}^d$  is complete ([6], Theorem 3.8),  $\{S_k x\}$  has a unique limit

$$Sx = \lim_{k \rightarrow \infty} S_k x = \lim_{k \rightarrow \infty} \sum_{r=0}^k (-1)^r C^r x = \sum_{r=0}^{\infty} (-1)^r C^r x.$$

It remains to show  $S = (I + C)^{-1}$ , that is,  $S(I + C)x = (I + C)Sx = x$  for each  $x$ . We have

$$\|S_k(I + C)x - x\| = \|S_k x + S_k Cx - x\| = \|(-1)^k C^{k+1} x\| \leq \gamma^{k+1} \|x\| \rightarrow 0,$$

as  $k \rightarrow \infty$ , so that  $\lim_{k \rightarrow \infty} S_k(I + C)x = x$  and  $S(I + C)x = x$ . Similar arguments give  $(I + C)Sx = x$ . That  $\sum_{r=0}^{\infty} (-1)^r C^r x$  is absolutely convergent follows from

$$\left\| \sum_{r=0}^{\infty} (-1)^r C^r x \right\| \leq \sum_{r=0}^{\infty} \|C^r x\| \leq \sum_{r=0}^{\infty} \gamma^r \|x\| = \frac{\|x\|}{1 - \gamma}$$

since  $\|x\|/(1 - \gamma)$  is finite.  $\blacktriangle$

**Example 8.21** Let  $U \in \mathbb{R}$  be the solution of (8.1) with  $d = 1$ ,  $A = (\bar{A} + R)$ ,  $\bar{A} = E[A]$ ,  $R$  a random variable, and  $B$  a constant. If  $|R/\bar{A}| \leq \gamma < 1$  a.s., the sequence  $\sum_{r=0}^k (-1)^r (R/\bar{A})^r (B/\bar{A})$  converges a.s. to  $U$  as  $k \rightarrow \infty$ , so that the solution of  $(A + R)U = B$  is  $U = \sum_{r=0}^{\infty} (-1)^r (R/\bar{A})^r (B/\bar{A})$ .  $\diamond$

*Proof* An alternative form of  $(\bar{A} + R)U = B$  is  $(1 + \bar{A}^{-1}R)U = \bar{A}^{-1}B$  so that  $U = (1 + \bar{A}^{-1}R)^{-1}(\bar{A}^{-1}B)$ . Let  $S_k(\omega) = 1 + \sum_{r=1}^k (-1)^r \bar{A}^{-r} R(\omega)^r$  be a sample of  $S_k$ . Since  $|R/\bar{A}| \leq \gamma < 1$  a.s., the numerical sequence

$$|S_k(\omega)| \leq 1 + \sum_{r=1}^k |\bar{A}^{-r} R(\omega)^r| \leq 1 + \sum_{r=1}^k \gamma^r = \frac{1 - \gamma^{k+1}}{1 - \gamma} \leq \frac{1}{1 - \gamma} < \infty$$

has a limit as  $k \rightarrow \infty$  for almost all  $\omega \in \Omega$ .  $\blacktriangle$

**Theorem 8.20** *If the entries of  $A$  in (8.1) have finite variance and  $\|\bar{A}^{-1}Rx\| \leq \gamma\|x\|$  a.s. for all  $x \in \mathbb{R}^d$ , then the Neumann series representation of the solution  $U$  of this equation is*

$$U = \sum_{r=0}^{\infty} U^{(r)}, \quad \text{where } U^{(r)} = -\bar{A}^{-1}RU^{(r-1)}, r = 1, 2, \dots, \quad \text{with } U^{(0)} = \bar{A}^{-1}B, \quad (8.59)$$

where  $\bar{A} = E[A]$  and  $R = A - \bar{A}$ , so that the mean and covariance matrices of  $U \simeq \tilde{U} = U^{(0)} + U^{(1)}$  are  $\mu_u = \bar{A}^{-1}B$  and  $\gamma_u = E[(\bar{A}^{-1}R\bar{A}^{-1}B)(\bar{A}^{-1}R\bar{A}^{-1}B)']$ .

*Proof* The solution  $U$  of (8.1) can be given in the form  $U = (\bar{A} + R)^{-1}B = (I + \bar{A}^{-1}R)^{-1}\bar{A}^{-1}B$ , so that  $U = \sum_{r=0}^{\infty} (-1)^r (\bar{A}^{-1}R)^r \bar{A}^{-1}B$  by (8.58). The approximate moments of  $U$  in (8.59) follow from the representation  $U \simeq U^{(0)} + U^{(1)}$  of  $U$  by straightforward calculations.  $\blacktriangle$

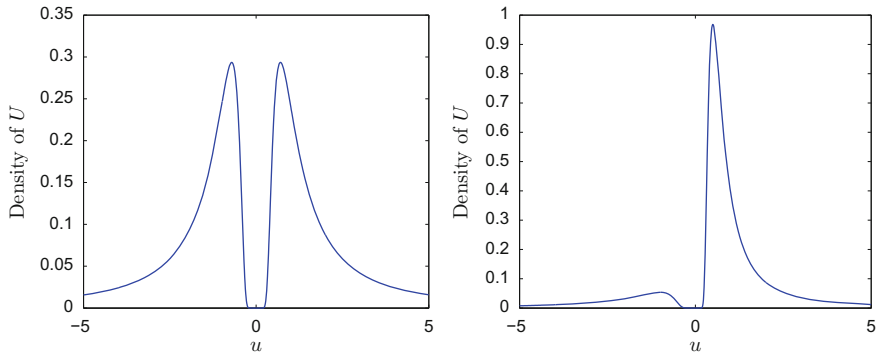
As for the perturbation and Taylor series methods, the assumption that  $B$  is deterministic is not restrictive. Moments and other properties of  $U$  can be calculated from truncations  $\tilde{U}^{(n)} = \sum_{r=0}^n (-1)^r (\bar{A}^{-1}R)^r \bar{A}^{-1}B$  of the Neumann series representation  $U = \sum_{r=0}^{\infty} (-1)^r (\bar{A}^{-1}R)^r \bar{A}^{-1}B$ . Since the functional form of  $\tilde{U}^{(n)}$  is available, its properties can be estimated efficiently by Monte Carlo simulation.

### 8.3.4 Equivalent Linearization

Let  $\tilde{U} = h(X)$  be a trial solution for (8.1) with  $A$  depending on  $X$  and  $B$  assumed to be deterministic, where  $h : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^d$  is a measurable function. Our objective is to find that  $h$  minimizes the error  $\|\tilde{U} - U\| = E[(\tilde{U} - U)'(\tilde{U} - U)^*]^{1/2}$ . This optimization problem cannot be solved since  $U$  is unknown. We consider an alternative optimization problem that selects a function  $h$  that minimizes the discrepancy  $\|Ah(X) - B\|$  between the left side of (8.1) with  $\tilde{U}$  in place of  $U$  and the right side of this equation.

**Definition 8.4** The equivalent linear solution of (8.1) has the form  $\tilde{U} = \alpha X + \beta$ , where  $\alpha$  and  $\beta$  are  $(d, d_x)$  and  $(d, 1)$  matrices with entries selected to minimize the mean square error  $\|A(\alpha X + \beta) - B\|$ .

*Example 8.22* Let  $d = 1$ ,  $B > 0$  a constant, and  $A = X$  a lognormal variable with mean  $\mu$ , variance  $\sigma^2$  and coefficient of variation  $v = \sigma/\mu$ . The equivalent linear



**Fig. 8.12** Densities of  $U$  defined by  $AU = 1$  for  $A \sim N(\mu, \sigma^2)$  with  $(\mu = 0, \sigma = 1)$  (left panel) and  $(\mu = 1, \sigma = 1)$  (right panel)

solution is  $\tilde{U} = \alpha X + \beta$  with  $\alpha = -B\mu^{-2}(1 + v^2)^{-4}$  and  $\beta = B\mu^{-1}(2 + v^2)(1 + v^2)^{-2}$  obtained by minimizing the objective function  $e = E[(B - \alpha X^2 - \beta X)^2]$ , that is, imposing the conditions  $\partial e / \partial \alpha = 0$  and  $\partial e / \partial \beta = 0$ . The approximate mean and variance of  $U$  are  $E[\tilde{U}] = B\mu^{-1}(1 + v^2)^{-4}(1 + 5v^2 + 4v^4 + v^6)$  and  $\text{Var}[\tilde{U}] = B^2\mu^{-2}(1 + v^2)^{-8}v^2$ .  $\diamond$

The linearization method may provide limited if any information on the probability law of  $U$  since  $\tilde{U}$  depends linearly on  $X$  while  $U$  is a nonlinear function of this random vector.

*Example 8.23* Consider the scalar equation  $AU = 1$  with  $A \sim N(\mu, \sigma^2)$ . Figure 8.12 shows densities of the solution  $U$  of  $AU = 1$  for two values of  $(\mu, \sigma)$ . These densities differ significantly from that of the solution  $\tilde{U} = \alpha A + \beta$  obtained by the equivalent linearization method, which is a Gaussian variable.  $\diamond$

*Proof* The distribution of  $U = 1/A$  is

$$P(U \leq u) = P(A \geq 1/u) = \Phi\left(\frac{\mu - 1/u}{\sigma}\right), \quad u > 0,$$

$$P(U \leq u) = P(A \in (1/u, 0)) = \frac{1}{2} - \Phi\left(\frac{1/u - \mu}{\sigma}\right), \quad u < 0,$$

so that its density is  $f_U(u) = (\sigma u^2)^{-1}[\phi((\mu - 1/u)/\sigma)1(u > 0) + \phi((1/u - \mu)/\sigma)1(u < 0)]$ ,  $u \in \mathbb{R}$ .  $\blacktriangle$

## 8.4 Exercises

**Exercise 8.1.** Show that the estimator  $\hat{F}_h$  in (8.9) is unbiased and its variance vanishes as  $n \rightarrow \infty$ .

**Exercise 8.2.** Repeat the calculations in Example 8.5 for a different choice of  $\{\tilde{x}^{(k)}\}$ ,  $k = 1, \dots, m$ , in  $[a, b]$ .



**Exercise 8.3.** Complete the proof of Theorem 8.4.

**Exercise 8.4.** Use the substitution principle in (8.15) to smooth the SROM-based distributions of the eigenvalues of matrix  $A$  in (8.16).

**Exercise 8.5.** Let  $U^{(n)}$ ,  $n = 1, 2, \dots$ , and  $U$  be  $\mathbb{R}^d$ -valued random variables in  $L^2(\Omega, \mathcal{F}, P)$ . Show that if  $U^{(n)}$  converges weakly to  $U$ , that is,  $\langle U^{(n)}, W \rangle \rightarrow \langle U, W \rangle$  as  $n \rightarrow \infty$ ,  $\forall W \in L^2(\Omega, \mathcal{F}, P)$ , then its weak limit  $U$  is unique.

**Exercise 8.6.** Show that in a finite dimensional Hilbert space  $H$  weak and strong convergence are equivalent concepts.

*Hint* Let  $(e^{(1)}, \dots, e^{(m)})$  be an orthonormal basis in  $H$  and let  $x_n$ ,  $n = 1, 2, \dots$ , be a sequence in  $H$  converging weakly to  $x \in H$ . Since  $x_n - x = \sum_{k=1}^m (\langle x_n, e^{(k)} \rangle - \langle x, e^{(k)} \rangle) e^{(k)}$ , we have  $\|x_n - x\|^2 = \sum_{k=1}^m |\langle x_n, e^{(k)} \rangle - \langle x, e^{(k)} \rangle|^2$ .

**Exercise 8.7.** Develop an optimization algorithm that delivers a SROM for a bivariate translation vector with coordinates  $X_1 \sim U(0, 1)$  and  $X_2 \sim EXP(\lambda)$ ,  $\lambda > 0$ , so that  $X_1 \stackrel{d}{=} \Phi(G_1)$ ,  $X_2 \stackrel{d}{=} -\ln(1 - \Phi(G_2))/\lambda$ , and  $(G_1, G_2)$  is a standard bivariate Gaussian vector with correlation coefficient  $\rho = E[G_1 G_2]$ .

**Exercise 8.8.** Find polynomial chaos expansions of increasing order  $n$  for  $X = |N(0, 1)|$ . Calculate and plot fourth order moments for  $X$  as a function of  $n$ . Comment on your findings.

**Exercise 8.9.** Repeat calculations in Example 8.17 by using Bernstein polynomials defined by (8.42), rather than Lagrange polynomials.

**Exercise 8.10.** Develop an optimization algorithm for calculating the probability  $P(g(X) \leq 0)$  by the approximation in (8.46), where  $X$  is a two-dimensional vector with independent  $N(0, 1)$  coordinates and  $g(x) = (x_1/a_1)^2 + (x_2/a_2)^2 - 1$  with  $(a_1, a_2) = (2, 1)$  and  $(a_1, a_2) = (4, 3)$ . Evaluate the accuracy of the resulting approximations by Monte Carlo simulation.

**Exercise 8.11.** Extend the approximate solution in (8.51) to the case in which  $A$  and  $B$  depend on  $X$ .

**Exercise 8.12.** Derive approximate expressions for the first two moments of  $U$  defined by (8.54). Assume the coordinates of  $N$  are  $N_i = U_i^2$ ,  $i = 1, \dots, d$ .

**Exercise 8.13.** Find the approximate second moments properties of  $U$  defined by  $AU = B$  with  $A$  and  $B$  in (8.52) by using the representation in (8.53).

**Exercise 8.14.** Solve the SAE in Example 8.19 by the perturbation and Neumann series methods.

**Exercise 8.15.** Complete the calculations in Example 8.22.

## References

1. Babuška IM, Chatzipantelidis P (2002) On solving elliptic stochastic partial differential equations. *Comput Methods Appl Mech Eng* 191:4093–4122
2. Babuška IM, Nobile F, Tempone R (2007) A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J Numer Anal* 45(3):1005–1034
3. Babuška IM, Strouboulis T (2001) The finite element method and its reliability. Clarendon Press, Oxford University Press, New York
4. Brabenec RL (1990) Introduction to real analysis. PWS-KENT Publishing Company, Boston
5. Breitung K (1984) Asymptotic approximation for multinormal integrals. *J Eng Mech ASCE* 110:357–366
6. Bryant V (1987) Metric spaces iteration and applications. Cambridge University Press, Cambridge
7. Cope JE, Rust BW (1979) Bounds on solutions of linear systems with inaccurate data. *SIAM J Numer Anal* 16(6):950–963
8. Ditlevsen O (1981) Uncertainty modeling with applications to multidimensional civil engineering systems. McGraw-Hill Inc., New York
9. Field RV, Grigoriu M (2007) Convergence properties of polynomial chaos approximations for  $L^2$  random variables. Technical report SAND2007-1262. Sandia National Laboratories, Albuquerque
10. Ghanem RG, Spanos PD (1991) Stochastic finite elements: a spectral approach. Springer, New York
11. Grigoriu M (2002) Stochastic calculus. Applications in science and engineering. Birkhäuser, Boston
12. Grünbaum FA (1975) The determinant of a random matrix. *Bull Am Math Soc* 81(2):446–448
13. Hairer E, Nørsett SP, Wanner G (1993) Solving ordinary differential equations I. Nonstiff problems. Springer, New York (Second revised edition)
14. Hildebrandt TH, Schoenberg IJ (1933) On linear functional operations and the moment problem for finite interval in one or several dimensions. *Ann Math* 34(2):317–328
15. Hinch EJ (1994) Perturbation methods. Cambridge University Press, Cambridge
16. Knight K (2000) Mathematical statistics. Chapman & Hall/CRC, New York
17. Kuo H-H (2006) Introduction to stochastic integration. Springer, New York
18. Loeven GIA, Witteveen JAS, Bijl H (2007) Probabilistic collocation: an efficient non-intrusive approach for arbitrary distributed parametric uncertainties. Reno, Nevada
19. Lorentz GG (1986) Bernstein polynomials. Chelsea Publishing Company, New York
20. Mehta ML (2004) Random matrices. Elsevier, Amsterdam
21. Nobile F, Tempone R, Webster CG (2008) An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J Numer Anal* 46(5):2411–2442
22. Powell MJD (1981) Approximation theory and methods. Cambridge University Press, Cambridge
23. Rubinstein R (1981) Simulation and the Monte Carlo method. Wiley, New York
24. Simmonds JG, Mann JE (1986) A first look at perturbation theory. Robert E. Krieger Publishing Company, Malabar
25. Söderling G (2006) The logarithmic norm. History and modern theory. *BIT Numer Math* 46:631–652
26. Sudret B, DerKiureghian A (2002) Comparisons of finite element reliability methods. *Probab Eng Mech* 17:337–348
27. Tricomi FG (1957) Integral equations. Dover Publications, Inc., New York

28. Wan X, Karniadakis GE (2005) An adaptive multi-element generalized polynomial chaos methods for stochastic differential equations. *J Comput Phys* 209(2):617–642. doi:[10.1016/j.jcp.2005.03.023](https://doi.org/10.1016/j.jcp.2005.03.023)C822
29. Xiu D, Karniadakis GE (2002) The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput* 24(2):619–644. doi:[10.1137/S1064827501387826](https://doi.org/10.1137/S1064827501387826)

# Chapter 9

## Stochastic Partial Differential Equations

### 9.1 Introduction

Deterministic partial differential equations (PDEs) have been used extensively in science and engineering to predict the behavior and assess the performance of a broad range of physical systems. These equations are likely to describe satisfactorily the behavior of a system at the macroscopic scale, but may provide insufficient, or even incorrect, information on the system state at the microscopic and mesoscopic scales [1]. At small scale, the coefficients of PDEs describing a system state fluctuate randomly in space. Probabilistic models, for example, random fields, need to be employed to characterize these coefficients [2–6]. We refer to the class of PDEs with random coefficients, source term, and/or end conditions as stochastic partial differential equations (SPDEs).

There are notable differences between the types of SPDEs in the mathematical and applied literature and between the methods used to solve these equations. In the applied literature, the random elements in the definition of SPDEs are usually colored random functions of space and time that may or may not have smooth samples. Weak forms of these equations are used to establish conditions for the existence and uniqueness of their solutions and construct numerical algorithms. The random elements in the definition of SPDEs in the mathematical literature are primarily white noise random functions of space and time. A generalized version of Itô's formula for semimartingales is the essential tool for analysis.

Our objectives are to establish conditions for the existence and uniqueness of the solutions of SPDEs and examine methods for solving these equations. It is assumed that both the functional form of the SPDEs and the probability law of the random elements in their definitions are known. The following two sections focus on SPDEs commonly considered in the mathematical literature. It is shown in [Sect. 9.3](#) that approximate versions of these equations obtained by space and time discretization can be solved by the methods in [Chaps. 7 and 8](#). [Sections 9.4 and 9.5](#) deal with SPDEs encountered in applications with random entries that have arbitrary and small uncertainty, respectively. The focus is on stochastic elliptic partial differential equations.

It is argued in [Sect. 9.4](#) that finite element and other numerical methods for deterministic PDEs [\[7, 8, 9, 10\]](#) cannot be extended directly to solve SPDEs. We examine the efficiency and accuracy of Monte Carlo, stochastic reduced order models, stochastic Galerkin, and stochastic collocation methods for solving SPDEs. [Section 9.5](#) reviews briefly the Taylor, perturbation, and Neumann series methods for solving SPDEs with coefficients of small uncertainty.

## 9.2 Stochastic Partial Differential Equations

Partial differential equations with random coefficients, source term, and/or initial and boundary conditions depending on space and time are referred to as stochastic partial differential equations (SPDEs). Generally, the random coefficients of the equations considered in the mathematical literature are white noise processes that are more general than those considered so far in the book since they depend on both time and space. The solution of these equations requires extensions of our previous developments related to stochastic integrals, differentiation formulas, and other advanced concepts on probability theory and random functions.

For example, let  $(\Omega, \mathcal{F}, P)$  be a probability space endowed with a filtration  $\mathcal{F}_t$ ,  $t \geq 0$ . A real-valued function  $\Phi(x, t)$ ,  $x \in \mathbb{R}^d$ ,  $t \geq 0$ , is said to be a spatially dependent martingale if it is an  $\mathcal{F}_t$ -martingale for each  $x \in \mathbb{R}^d$ . If  $\Sigma(x, t)$  is an  $\mathcal{F}_t^B$ -adapted, real-valued process on  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$  such that  $\int_0^\tau |\Sigma(x, t)|^2 dt < \infty$  a.s. for each  $x \in \mathbb{R}^d$  and  $\mathcal{F}_t^B \subset \mathcal{F}_t$ , then the Itô integrals  $I(x) = \int_0^\tau \Sigma(x, t) dB(t)$  are well defined for each  $x \in \mathbb{R}^d$ , where  $\mathcal{F}_t^B = \sigma(B(s), 0 \leq s \leq t)$  denotes the filtration generated by a Brownian motion  $B$  defined on  $(\Omega, \mathcal{F}, P)$  ([Sect. 4.4](#)). However,  $I(x)$ ,  $x \in \mathbb{R}^d$ , may not be a random field since the union of the null sets  $N(x)$  on which  $I(x)$  is not defined may not be a null set or even a measurable set. Also, the Itô formulas established in [Sects. 5.2](#) and [5.3](#) need to be generalized to be applied to functions of spatially dependent martingales and semimartingales ([\[11\]](#), Theorem 2.3, p. 20). The reader interested in mathematical developments on SPDEs is referred to [\[11, 12\]](#). Our discussion in this section is based on [\[11\]](#).

Let  $Z(x)$ ,  $x \in D$ , be a real-valued random field with mean 0, finite variance, and correlation function  $r(x, y) = E[Z(x)Z(y)]$  assumed to be continuous, where  $D$  is a bounded subset of  $\mathbb{R}^d$ . Since  $Z(x)$  has finite variance,  $\int_{D^2} r(x, y)^2 dx dy < \infty$  and

$$Z(x) = \sum_{k=1}^{\infty} \sqrt{\mu_k} Z_k \varphi_k(x), \quad x \in D, \quad (9.1)$$

holds in the mean square sense, where  $\{\mu_k \geq 0\}$  and  $\{\varphi_k\}$  denote the eigenvalues and the eigenfunctions of the integral equation  $\int_D r(x, y) \varphi(y) dy = \mu \varphi(x)$ ,  $x \in D$ , and  $\{Z_k\}$  are uncorrelated random variables with mean 0 and variance 1. If  $Z(x)$  is a Gaussian field, then  $\{Z_k\}$  are independent  $N(0, 1)$  variables ([Sect. 3.6.5](#)).

The correlation functions  $r_n(x, y) = E[Z^{(n)}(x) Z^{(n)}(y)] = \sum_{k=1}^n \mu_k \varphi_k(x) \varphi_k(y)$  of the random fields  $Z^{(n)}(x) = \sum_{k=1}^n \sqrt{\mu_k} Z_k \varphi_k(x)$ ,  $n = 1, 2, \dots$ , obtained by truncating the representation of  $Z(x)$  after  $n$  terms converge to  $r(x, y) = E[Z(x) Z(y)]$  as  $n \rightarrow \infty$  (Theorem 3.22). It can be assumed that the eigenfunctions in (9.1) form an orthonormal sequence, that is,  $\int_D \varphi_k(x) \varphi_l(x) dx = \delta_{kl}$ .

**Definition 9.1** Let  $\{B_k\}$  be independent Brownian motions defined on a probability space  $(\Omega, \mathcal{F}, P)$ , and set

$$W^{(n)}(x, t) = \sum_{k=1}^n \sqrt{\mu_k} \varphi_k(x) B_k(t), \quad x \in D, \quad t \in [0, \tau], \quad (9.2)$$

with  $(\mu_k, \varphi_k)$  as in (9.1). We refer to  $\{W^{(n)}\}$  as spatially dependent stochastic processes.

The processes  $W^{(n)}(x, t)$  in (9.2) have zero means and correlation functions

$$E[W^{(n)}(x, t) W^{(n)}(y, s)] = \sum_{k=1}^n \mu_k \varphi_k(x) \varphi_k(y) (s \wedge t) = r_n(x, y) (s \wedge t), \quad n = 1, 2, \dots, \quad (9.3)$$

so that  $r_n(x, y) \rightarrow r(x, y)$  by properties of Karhunen–Loève representation. Also, for  $n > m$ ,  $\|W^{(n)} - W^{(m)}\|_{L^2(D)}^2 = \sum_{k=m+1}^n \mu_k B_k(t)^2 \int_D \varphi_k(x)^2 dx = \sum_{k=m+1}^n \mu_k B_k(t)^2$  by properties of  $\{\varphi_k(x)\}$ , so that  $E[\|W^{(n)} - W^{(m)}\|_{L^2(D)}^2] = t \sum_{k=m+1}^n \mu_k \leq \tau \sum_{k=m+1}^n \mu_k$ . Under the assumption  $\sum_{k=1}^\infty \mu_k < \infty$ , we have  $\sum_{k=m+1}^n \mu_k \rightarrow 0$  as  $m, n \rightarrow \infty$ . Since  $\|W^{(n)}(\cdot, t) - W^{(m)}(\cdot, t)\|_{L^2(D)}^2$  is a real-valued continuous submartingale, there exists a constant  $c > 0$  such that ([11], Theorem 3.3, p. 11)

$$\begin{aligned} & E\left[\sup_{0 \leq t \leq \tau} \|W^{(n)}(\cdot, t) - W^{(m)}(\cdot, t)\|_{L^2(D)}^2\right] \\ & \leq c E[\|W^{(n)}(\cdot, \tau) - W^{(m)}(\cdot, \tau)\|_{L^2(D)}^2] \leq c \tau \sum_{k=m+1}^n \mu_k, \end{aligned}$$

which implies  $E[\sup_{0 \leq t \leq \tau} \|W^{(n)}(\cdot, t) - W^{(m)}(\cdot, t)\|_{L^2(D)}^2] \rightarrow 0$  as  $m, n \rightarrow \infty$ . Hence,  $\{W^{(n)}(x, \cdot)\}$  are Brownian processes for each  $x \in D$  and

$$W(x, t) = \lim_{n \rightarrow \infty} W^{(n)}(x, t) = \sum_{k=1}^\infty \sqrt{\mu_k} \varphi_k(x) B_k(t), \quad x \in D, \quad t \in [0, \tau], \quad (9.4)$$

exists in the mean square sense.

**Definition 9.2** The formal time derivative  $\dot{W}(x, t) = \partial W(x, t)/\partial t$  of  $W(x, t)$  is called spatially dependent Gaussian white noise. We also refer to the temporal derivatives  $\dot{W}^{(n)}(x, t) = \partial W^{(n)}(x, t)/\partial t$  as spatially dependent white noise processes. Formal calculations show that the white noise processes  $\dot{W}(x, t)$  and  $\dot{W}^{(n)}(x, t)$

have mean 0. Their covariance functions are  $E[\dot{W}(x, t) \dot{W}(y, s)] = r(x, y) \delta(s - t)$  and  $E[\dot{W}^{(n)}(x, t) \dot{W}^{(n)}(y, s)] = r_n(x, y) \delta(s - t)$ , respectively.

The following example adapted from [11] (Sect. 3.3) illustrates technical difficulties involved in the solution of SPDEs of the type considered in the mathematical literature, and the need to introduce simplifying assumptions, that may be too restrictive for applications. The problem in this example is also solved in Example 9.2 by using finite differences and methods in Sect. 7.2 to approximate spatial derivatives and solve the resulting ordinary differential equations driven by random noise. The approach in Example 9.2 is conceptually simple, computationally efficient, and accurate.

*Example 9.1* Let  $U(x, t)$  be a real-valued random function satisfying the stochastic partial differential equation

$$\frac{\partial U(x, t)}{\partial t} = \left( \beta \frac{\partial^2}{\partial x^2} - \alpha \right) U(x, t) + \dot{W}(x, t), \quad x \in (0, l), \quad t \in (0, \tau) \quad (9.5)$$

with boundary and initial conditions  $U(0, t) = U(l, t) = 0$  and  $U(x, 0) = h(x)$ , where  $\alpha, \beta, l > 0$  are constants,  $h(x)$  is a smooth deterministic function and  $\dot{W}(x, t)$  denotes a spatially dependent Gaussian white noise. The solution of (9.5) can be given in the form

$$U(x, t) = \sum_{k=1}^{\infty} U_k(t) e_k(x), \quad x \in (0, l), \quad t \in (0, \tau), \quad (9.6)$$

where  $U_k(t)$  are Ornstein–Uhlenbeck processes defined by

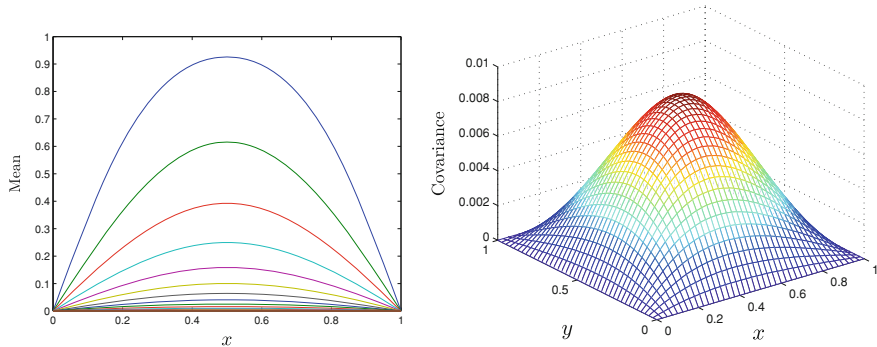
$$dU_k(t) = -\lambda_k U_k(t) dt + \sqrt{\mu_k} dB_k(t), \quad t \in (0, \tau), \quad (9.7)$$

$\{B_k(t)\}$  are independent Brownian motions, the initial state for (9.7) is  $U_k(0) = \langle h, e_k \rangle / \|e_k\|^2 = 2(1 - \cos(\rho_k l)) / \rho_k$ ,  $\|\cdot\|$  is the norm in  $L^2(0, l)$ , and  $\lambda_k$  and  $e_k(x)$  denote the eigenvalues and eigenfunction of the operator  $\mathcal{A} = \beta \partial^2 / \partial x^2 - \alpha$ , so that they satisfy  $-\beta e''(x) + \alpha e(x) = \lambda e(x)$  or, equivalently,  $e''(x) + \rho^2 e(x) = 0$ , with the boundary conditions  $e(0) = e(l) = 0$ , where  $\rho^2 = (\lambda - \alpha) / \beta$ . We have  $\rho_k = k\pi / l$  and  $e_k(x) = \sin(k\pi x / l)$ ,  $k = 1, 2, \dots$ , so that  $\lambda_k = \alpha + \beta \rho_k^2$ . Since

$$U_k(t) = U_k(0) e^{-\lambda_k t} + \sqrt{\mu_k} I_k(t), \quad \text{where } I_k(t) = \int_0^t e^{-\lambda_k(t-s)} dB_k(s), \quad (9.8)$$

we have

$$U(x, t) = \sum_{k=1}^{\infty} U_k(0) e^{-\lambda_k t} e_k(x) + \sum_{k=1}^{\infty} \sqrt{\mu_k} I_k(t) e_k(x). \quad (9.9)$$



**Fig. 9.1** Expectations  $E[U^{(n)}(x, t)]$  as functions of  $x \in (0, l)$  at several times (*left panel*) and covariance matrix of  $U^{(n)}(x, \tau)$  (*right panel*)

Denote by

$$U^{(n)}(x, t) = \sum_{k=1}^n U_k(0) e^{-\lambda_k t} e_k(x) + \sum_{k=1}^n \sqrt{\mu_k} I_k(t) e_k(x) \quad (9.10)$$

the series representation of  $U(x, t)$  truncated after the first  $n$  terms.

The mean and correlation functions of  $U(x, t)$  are (Sect. 7.2.2)

$$\begin{aligned} E[U(x, t)] &= \sum_{k=1}^{\infty} 2 \frac{1 - \cos(\rho_k \ell)}{\rho_k} e^{-\lambda_k t} e_k(x) \quad \text{and} \\ E[U(x, t) U(y, s)] &= \sum_{k=1}^{\infty} \frac{\mu_k}{2 \lambda_k} \left( e^{-\lambda_k |s-t|} - e^{\lambda_k (s+t)} \right) e_k(x) e_k(y). \end{aligned} \quad (9.11)$$

Numerical results have been obtained for  $\alpha = \beta = 1$ ,  $l = 1$ ,  $\tau = 1$ ,  $h(x)=1$ , and  $U^{(n)}(x, t) = \sum_{k=1}^n U_k(t) e_k(x)$  with  $n=30$ . The left panel in Fig. 9.1 shows expectations of  $U^{(n)}(x, t)$  as functions of  $x \in (0, l)$  at several times. The expectation of  $U(x, t)$  approaches 0 as time increases. The covariance matrix of  $U^{(n)}(x, \tau)$  shown in the right panel of the figure is zero on the boundary of  $(0, l)^2$  in agreement with the specified boundary values for  $U(x, t)$ .

The deterministic part  $U^{(d,n)} = \sum_{k=1}^n U_k(0) e^{-\lambda_k t} e_k(x)$  of  $U^{(n)}$  in (9.10) converges to that of  $U$  by properties of deterministic partial differential equations. The random part  $U^{(r,n)}(x, t) = \sum_{k=1}^n \sqrt{\mu_k} I_k(t) e_k(x)$  of  $U^{(n)}$  converges to that of  $U$  in mean square provided  $\sum_{k=1}^{\infty} \mu_k < \infty$ . Moreover,  $U(x, t)$  is mean square continuous and satisfies (9.5) in the weak sense, that is,

$$\langle U(\cdot, t), \phi \rangle = \langle h, \phi \rangle + \int_0^t \langle U(\cdot, s) \mathcal{A}[\phi] \rangle ds + \langle W(\cdot, t), \phi \rangle \quad \text{a.s.}, \quad (9.12)$$

where  $\phi \in C_0^\infty(0, l)$  are trial functions.  $\diamond$



*Proof* The solution of the homogeneous version of (9.5) by separation of variables has the form  $U(x, t) = X(x) T(t)$  so that  $X(x) \dot{T}(t) = \beta X''(x) T(t) - \alpha X(x) T(t)$  implying  $\dot{T}(t)/T(t) = (\beta X''(x) - \alpha X(x))/X(x) = -\lambda$ , where  $\lambda > 0$  is a constant. The general homogeneous solution is  $U(x, t) = \sum_{k=1}^{\infty} c_k e^{-\lambda_k t} e_k(x)$ , where  $c_k$  are constants and  $\lambda_k$  and  $e_k(x)$  denote the eigenvalues and eigenfunctions of the differential equation  $\beta e''(x) - \alpha e(x) = -\lambda e(x)$  or  $e''(x) + \rho^2 e(x) = 0$  satisfying the boundary conditions  $e(0) = e(l) = 0$ . The notation  $\rho^2 = (\lambda - \alpha)/\beta$  is meaningful since  $0 < \alpha < \lambda_1 \leq \lambda_2 \leq \dots$  and  $\beta > 0$ .

If the white noise admits the representation  $\dot{W}(x, t) = \sum_{k=1}^{\infty} \sqrt{\mu_k} e_k(x) \dot{B}_k(t)$ , that is,  $\varphi_k(x)$  in (9.2) are the eigenfunctions  $e_k(x)$  of  $\mathcal{A} = \beta \partial^2 / \partial x^2 - \alpha$  and  $\dot{B}_k(t)$  denotes the formal derivative of  $B_k(t)$ , then (9.5) becomes

$$\sum_{k \geq 1} [\dot{U}_k(t) + \lambda_k U_k(t) - \sqrt{\mu_k} \dot{B}_k(t)] e_k(x) = 0,$$

which gives (9.7) since this equality must hold for every  $x \in (0, l)$ .

The mean square convergence of the random part  $U^{(r,n)}$  of  $U^{(n)}$  to the random part  $U^{(r)}$  of  $U$  holds under the assumption  $\sum_{k=1}^{\infty} \mu_k < \infty$  since

$$\begin{aligned} E[\|U^{(r,n)}(\cdot, t)\|^2] &= \frac{l}{2} \sum_{k=1}^n \mu_k E[I_k(t)^2] = \frac{l}{4} \sum_{k=1}^n \frac{\mu_k}{\lambda_k} (1 - e^{-2\lambda_k t}) \leq \frac{l}{4} \sum_{k=1}^n \frac{\mu_k}{\lambda_k} \\ &\leq \frac{l}{4\alpha} \sum_{k=1}^n \mu_k \leq \frac{l}{4\alpha} \sum_{k=1}^{\infty} \mu_k < \infty \end{aligned}$$

so that  $\sup_{0 \leq t \leq \tau} E[\|U^{(r,n)}(\cdot, t)\|^2] < \infty$  and  $\sup_{0 \leq t \leq \tau} E[\|U^{(r)}(\cdot, t) - U^{(r,n)}(\cdot, t)\|^2] \leq (l/(4\alpha)) \sum_{k=n+1}^{\infty} \mu_k \rightarrow 0$  as  $n \rightarrow \infty$  since  $\sum_{k=1}^{\infty} \mu_k < \infty$ . Similar arguments can be used to show that  $U^{(r)}$  is mean square continuous, that is,  $E[\|U^{(r)}(\cdot, s) - U^{(r)}(\cdot, t)\|^2] \rightarrow 0$  as  $|s - t| \rightarrow 0$  (Exercise 9.1).

That  $U^{(n)}(x, t)$  in (9.6) is a weak solution for (9.5), that is,  $U(x, t)$  satisfies (9.12), follows from the formula  $U^{(n)}(x, t) = h(x) + \int_0^t \mathcal{A}[U^{(n)}(x, s)] ds + W(x, t)$  resulting from (9.5) by time integration, the representation of  $U^{(n)}(x, t)$  in (9.10), and the equalities

$$\begin{aligned} \langle U^{(n)}(\cdot, t), \phi \rangle &= \langle h^{(n)}, \phi \rangle - \sum_{k=1}^n \lambda_k \int_0^t U_k(s) ds \langle e_k, \phi \rangle + \langle W^{(n)}(\cdot, t), \phi \rangle \\ &= \langle h^{(n)}, \phi \rangle + \sum_{k=1}^n \int_0^t U_k(s) ds \langle \mathcal{A}[e_k], \phi \rangle + \langle W^{(n)}(\cdot, t), \phi \rangle \\ &= \langle h^{(n)}, \phi \rangle + \sum_{k=1}^n \int_0^t U_k(s) ds \langle e_k, \mathcal{A}[\phi] \rangle + \langle W^{(n)}(\cdot, t), \phi \rangle \\ &= \langle h^{(n)}, \phi \rangle + \int_0^t \langle U^{(n)}(\cdot, s), \mathcal{A}[\phi] \rangle ds + \langle W^{(n)}(\cdot, t), \phi \rangle, \end{aligned}$$

where  $h^{(n)} = \sum_{k=1}^n \langle h, e_k \rangle e_k$ ,  $\phi$  is a test function with the properties  $\phi \in C^\infty[0, l]$ , and  $\phi(0) = \phi(l) = 0$ . In the above equalities, we use the fact that  $\mathcal{A}[e_k] = \beta e_k'' - \alpha e_k$  is an adjoint operator (Sect. B.4.4).

The weak solution is unique since, if  $U^{(n)}(x, t) = \sum_{k=1}^n U_k(t) e_k(x)$  and  $\tilde{U}^{(n)}(x, t) = \sum_{k=1}^n \tilde{U}_k(t) e_k(x)$  are two weak solutions, then

$$V(x, t) = U^{(n)}(x, t) - \tilde{U}^{(n)}(x, t) = \sum_{k=1}^n (U_k(t) - \tilde{U}_k(t)) e_k(x) = \sum_{k=1}^n V_k(t) e_k(x),$$

so that  $V_k(t) = -\lambda_k \int_0^t V_k(s) ds$  for  $\phi = e_k$  and

$$|V_k(t)|^2 = \lambda_k^2 \left( \int_0^t V_k(s) ds \right)^2 \leq \lambda_k^2 t \int_0^t V_k(s)^2 ds$$

by the Cauchy–Schwarz inequality. This bound on  $|V_k(t)|^2$  and Gronwall's inequality imply  $V_k(t) = 0$ ,  $k = 1, \dots, n$ , a.s. for all  $t \in [0, \tau]$  (Exercise 9.2).  $\blacktriangle$

The approach in Example 9.1 for solving stochastic partial differential equations is not common in the applied literature. The following two-step approach is usually preferred in applications. First, partial differential equations are approximated by ordinary differential equations by space discretization or difference equations by space-time discretization. Second, methods of random vibration (Sect. 7.2) are used to solve the approximate equations constructed in the previous step.

*Example 9.2* Consider the stochastic partial differential equation in Example 9.1. Let  $0 = x_0 < x_1 < \dots < x_i < \dots < x_n < x_{n+1} = l$  be equally spaced points in  $(0, l)$ ,  $\Delta x = x_i - x_{i-1}$ ,  $i = 1, \dots, n+1$ , and  $\Delta t > 0$ . The finite difference version of (9.5) is

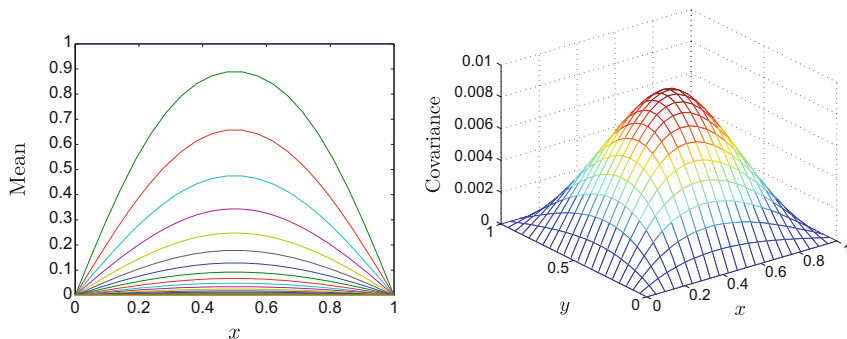
$$\begin{aligned} Y_i(t + \Delta t) &= a Y_{i-1}(t) + b Y_i(t) + a Y_{i+1}(t) + \dot{W}_i(t) \Delta t, \quad \text{for } i = 2, \dots, n-1, \\ Y_1(t + \Delta t) &= b Y_1(t) + a Y_2(t) + \dot{W}_1(t) \Delta t, \quad \text{for } i = 1, \\ Y_n(t + \Delta t) &= a Y_{n-1}(t) + b Y_n(t) + \dot{W}_n(t) \Delta t, \quad \text{for } i = n, \end{aligned} \quad (9.13)$$

where  $Y_i(t) = U(x_i, t)$ ,  $\dot{W}_i(t) = \dot{W}(x_i, t)$ ,  $a = \beta \Delta t / (\Delta x)^2$ ,  $b = 1 - 2\beta \Delta t / (\Delta x)^2 - \alpha \Delta t$ , and  $t \geq 0$ . The finite difference equations for  $i = 1$  and  $i = n$  account for the boundary values of  $U(x, t)$ . The matrix form of (9.13) is

$$Y(t + \Delta t) = A Y(t) + L N, \quad t \geq 0, \quad (9.14)$$

where  $Y(t) = (Y_1(t), \dots, Y_n(t))'$ ,  $A$  is an  $(n, n)$ -matrix with zero entries except for  $A(i, i) = b$  and  $A(i, i-1) = A(i, i+1) = a$ ,  $L$  denotes a lower triangular matrix such that  $L L' = R = \{r(x_i, x_j)\}$ ,  $i, j = 1, \dots, n$ , and  $N$  is an  $\mathbb{R}^n$ -valued random variable with independent  $N(0, \Delta t)$  coordinates.

The left and right panels in Fig. 9.2 are analogues to those in Fig. 9.1. They give the expectations of  $Y(t)$  at various times and the covariance of  $Y(\tau)$  for the same



**Fig. 9.2** Expectations of  $Y(t)$  at several times (*left panel*) and covariance matrix of  $Y(\tau)$  (*right panel*)

values of  $\alpha$ ,  $\beta$ ,  $l$ , and  $h(x)$  as in Example 9.1,  $n = 20$ , and  $\Delta t = 0.001$ . The solution of (9.5) by finite differences is conceptually simple, provides results similar to those in Example 9.1, and uses familiar concepts. Moreover, the finite difference method is not restricted to noise processes  $W(x, t)$  with the spatial and temporal correlations in Example 9.1.  $\diamond$

*Proof* The finite difference formulas in Theorem 7.1 applied to (9.14) give  $\mu(t + \Delta t) = A \mu(t)$  with  $\mu(0) = (h(x_1), \dots, h(x_n))'$  and  $\gamma(t + \Delta t) = A \gamma(t) A' + R \Delta t$  with  $\gamma(0) = 0$ , where  $\mu(t) = E[Y(t)]$  and  $\gamma(t) = E[(Y(t) - \mu(t))(Y(t) - \mu(t))']$ . The time and space steps,  $\Delta t$  and  $\Delta x$ , have to be selected such that the eigenvalues of  $A$  are included in the unit disc centered at the origin of the complex plane.  $\blacktriangle$

### 9.3 Discrete Approximations of SPDEs

It was shown in Example 9.2 that finite difference versions of SPDEs can be used to solve these equations approximately. Consider a SPDE  $\mathcal{L}[U(x, t)] = V(x, t)$  defined on an open bounded subset  $D$  of  $\mathbb{R}^d$  in a time interval  $(0, \tau)$ , and let  $\mathcal{L}_k^n[U_k^n] = V_k^n$  be a finite difference version of this equation defined on a space–time lattice or mesh in  $D \times (0, \tau)$ , where  $\{U_k^n\}$  are finite difference approximations for  $\{U(x_k, t_n)\}$ ,  $V_k^n = V(x_k, t_n)$ ,  $x_k$  is the coordinate of node  $k$ ,  $t_n = n \Delta t$ , and  $\Delta t > 0$  denotes the time step. Our objective is to establish conditions under which  $\{U_k^n\}$  provides an approximation for  $U(x, t)$ ,  $(x, t) \in D \times (0, \tau)$ . It is assumed that  $\mathcal{L}[U(x, t)] = V(x, t)$  is well-posed, that is, the magnitude of changes in  $U(x, t)$  caused by changes in data can be bounded by the magnitude of changes in data scaled by a constant.

Let  $U(x, t)$  be the potential in a random heterogeneous material specimen occupying an open bounded subset  $D$  of  $\mathbb{R}^d$ . Suppose the material diffusivity/conductivity in this specimen can be modeled by a real-valued random field  $\Sigma(x)$ ,  $x \in D$ , defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Then  $U(x, t)$  is the solution of

$$\frac{\partial U(x, t)}{\partial t} = \sum_{p=1}^d \frac{\partial \Sigma(x)}{\partial x_p} \frac{\partial U(x, t)}{\partial x_p} + \Sigma(x) \Delta U(x, t) + V(x, t), \quad x \in D, \quad t \geq 0, \quad (9.15)$$

accompanied by initial and boundary conditions, where  $\Delta = \sum_{p=1}^d \partial^2 / \partial x_p^2$  denotes the Laplace operator and  $V(x, t)$  is the source term at location  $x \in D$  and time  $t \geq 0$ . An alternative form of (9.15) is  $\mathcal{L}[U(x, t)] = V(x, t)$  with  $\mathcal{L}[U(x, t)] = \partial U(x, t) / \partial t - \nabla \cdot (\Sigma(x) \nabla U(x, t))$ .

It is assumed that (1) almost all the samples of the conductivity field  $\Sigma$  have continuous first order derivatives and take values in a bounded interval  $[a, b]$ ,  $0 < a < b < \infty$ , (2) the source term  $V$  is a deterministic function that is continuous in its arguments, and (3) the initial and the boundary conditions are deterministic and provide adequate information on  $U$  in  $D$  at the initial time  $t = 0$  and on the boundary  $\partial D$  of  $D$  at all times  $t > 0$  such that (9.15) has a unique solution for almost all samples of  $\Sigma(x)$ . Since  $\Sigma(x)$  is a random field, (9.15) is a stochastic partial differential equation and its solution is a random function of time and space defined on the same probability space as the conductivity random field.

We use finite differences to construct an ordinary differential equation with state  $Y(t)$ ,  $t \geq 0$ , approximating the solution  $U(x, t)$  of (9.15). The case  $d = 1$  is considered in details. The extension to the case  $d > 1$  is also discussed. Let  $U(x, t)$  be the solution of (9.15) with  $d = 1$  and  $D = (0, l)$ ,  $0 < l < \infty$ . For simplicity,  $V(x, t)$  is assumed to be a specified deterministic function, and  $U(x, t)$  is required to satisfy the deterministic initial and boundary conditions  $U(x, 0) = f(x)$  for  $x \in D$ , and  $U(0, t) = g_1(t)$  and  $U(l, t) = g_2(t)$  for  $t \geq 0$ .

Let  $h = l/(v + 1)$  and  $\Delta t > 0$  denote space and time steps, where  $v \geq 0$  is an integer. The finite difference version of (9.15) is

$$\begin{aligned} U^*(x, t + \Delta t) = & \frac{\Delta t}{h^2} A_1(x) U^*(x + h, t) + \left(1 - \frac{\Delta t}{h^2} A_2(x)\right) U^*(x, t) \\ & + \frac{\Delta t}{h^2} A_3(x) U^*(x - h, t) + V(x, t) \Delta t, \end{aligned} \quad (9.16)$$

where the coefficients  $A_1(x) = (\Sigma(x + h) - \Sigma(x - h) + 4 \Sigma(x))/4$ ,  $A_2(x) = 2 \Sigma(x)$ , and  $A_3(x) = (-\Sigma(x + h) + \Sigma(x - h) + 4 \Sigma(x))/4$  depend on the conductivity random field and  $U^*$  is defined at the nodes of the finite difference mesh with spatial and temporal coordinates  $x_k = k h$  and  $t_n = n \Delta t$ , that is,  $U^*(x_k, t_n) = U_k^n$  and  $\mathcal{L}_k^n[U_k^n] = V_k^n$  with the previous notations. The finite difference operator  $\mathcal{L}_k^n$  in (9.16) has been obtained from  $\mathcal{L}$  in (9.15) by approximating the spatial and temporal derivatives of  $U(x, t)$  by central and forward finite differences, respectively.

**Theorem 9.1** *Let  $e(x, t) = U(x, t) - U^*(x, t)$  be the discrepancy between the solutions  $U$  and  $U^*$ , and let  $\mathcal{E}(t) = \max_{0 \leq x \leq l} |e(x, t)|$ . If  $U$  is six times differentiable in  $x$  and three times differentiable in  $t$ ,  $A_1(x), A_3(x) > 0$ ,  $x \in [0, l]$ , for almost all samples of  $\Sigma(x)$ , and  $\Sigma(x)$  takes values in a bounded interval  $[a, b]$ ,  $0 < a < b < \infty$ , almost surely, then*

$$\frac{d\mathcal{E}(t)}{dt} \leq c_1(h), \quad \text{with } \mathcal{E}(0) = 0 \quad \text{and } c_1(h) > 0, \quad (9.17)$$

so that  $\mathcal{E}(t) \leq c_1(h)t$ . Since  $\lim_{h \rightarrow 0} c_1(h) = 0$ , the error  $\mathcal{E}(t)$  can be made as small as desired by a proper selection of  $h$ .

*Proof* Our arguments extend slightly upon considerations in [13] (Sect. 52) on deterministic parabolic differential equations. Since  $U(x, t)$  does not satisfy (9.16), we have

$$\begin{aligned} U(x, t + \Delta t) = & \frac{\Delta t}{h^2} A_1(x) U(x + h, t) + \left(1 - \frac{\Delta t}{h^2} A_2(x)\right) U(x, t) \\ & + \frac{\Delta t}{h^2} A_3(x) U(x - h, t) + V(x, t) \Delta t + T(x, t), \end{aligned} \quad (9.18)$$

where  $T(x, t)$  is a non-zero function. The error  $e(x, t) = U(x, t) - U^*(x, t)$  is the solution of the finite difference equation

$$\begin{aligned} e(x, t + \Delta t) = & \frac{\Delta t}{h^2} A_1(x) e(x + h, t) + \left(1 - \frac{\Delta t}{h^2} A_2(x)\right) e(x, t) \\ & + \frac{\Delta t}{h^2} A_3(x) e(x - h, t) + T(x, t), \end{aligned} \quad (9.19)$$

with zero initial and boundary conditions since  $U$  and  $U^*$  satisfy the same initial and boundary conditions. Since almost all samples of  $\Sigma$  take values in a bounded interval  $[a, b]$ ,  $0 < a < b < \infty$ , the functions  $A_1, A_2, A_3$  are bounded a.s. and

$$\begin{aligned} |e(x, t + \Delta t)| \leq & \frac{\Delta t}{h^2} |A_1(x)| |e(x + h, t)| + \left|1 - \frac{\Delta t}{h^2} A_2(x)\right| |e(x, t)| \\ & + \frac{\Delta t}{h^2} |A_3(x)| |e(x - h, t)| + |T(x, t)| \\ \leq & \left[ \frac{\Delta t}{h^2} |A_1(x)| + \left|1 - \frac{\Delta t}{h^2} A_2(x)\right| + \frac{\Delta t}{h^2} |A_3(x)| \right] \mathcal{E}(t) + |T(x, t)|, \end{aligned} \quad (9.20)$$

holds for almost all samples of  $\Sigma(x)$ . We assume that the conditions  $A_1(x), A_3(x) \geq 0$  and  $1 - (\Delta t/h^2) A_2(x) \geq 0$  hold a.s., so that the square bracket on the right side of (9.20) is equal to 1 and

$$|e(x, t + \Delta t)| \leq \mathcal{E}(t) + \max_{0 \leq x \leq l, t \geq 0} |T(x, t)| \leq \mathcal{E}(t) + M, \quad (9.21)$$

where  $M$  is an upper bound on  $|T(x, t)|$  for  $(x, t) \in (0, l) \times (0, \tau)$ . The inequality (9.21) implies  $\mathcal{E}(t + \Delta t) \leq \mathcal{E}(t) + M$ , so that  $\mathcal{E}(n \Delta t) \leq n M$ ,  $n = 0, 1, \dots$  since  $\mathcal{E}(0) = 0$ . The assumptions  $A_1(x), A_3(x) \geq 0$ , and  $1 - (\Delta t/h^2) A_2(x) \geq 0$  hold a.s. for  $5a - b \geq 0$  and proper selection of  $h$  and  $\Delta t$  since  $A_2(x)$  is bounded. The condition  $5a - b \geq 0$  is very strong, that is,  $A_1(x), A_3(x)$  may be positive although  $5a - b \geq 0$  is not satisfied, for example, the case in which  $\Sigma(x)$  has continuous samples and  $h$  is sufficiently small.

We now construct an upper bound  $M$  on  $|T(x, t)|$ . To simplify our discussion, the error related to the approximation  $\Sigma'(x) \simeq (\Sigma(x + h) - \Sigma(x - h))/(2h)$  is not considered. Under the assumptions that the solution  $U(x, t)$  is six and three times differentiable in  $x$  and  $t$ , respectively, the following Taylor expansions of  $U(x, t)$  hold

$$\begin{aligned}\frac{\partial U(x, t)}{\partial t} &= \frac{U(x, t + \Delta t) - U(x, t)}{\Delta t} - \frac{\Delta t}{2} \frac{\partial^2 U(x, t)}{\partial t^2} - \frac{\Delta t^2}{6} \frac{\partial^3 U(x, \tau)}{\partial t^3} \\ \frac{\partial U(x, t)}{\partial x} &= \frac{U(x + h, t) - U(x - h, t)}{2h} - \frac{h^2}{6} \frac{\partial^3 U(x, t)}{\partial x^3} - \frac{h^4}{120} \frac{\partial^5 U(\xi, t)}{\partial x^5} \\ \frac{\partial^2 U(x, t)}{\partial x^2} &= \frac{U(x + h, t) - 2U(x, t) + U(x - h, t)}{h^2} - \frac{h^2}{12} \frac{\partial^4 U(x, t)}{\partial x^4} - \frac{h^4}{360} \frac{\partial^6 U(\xi, t)}{\partial x^6}\end{aligned}\quad (9.22)$$

where  $\tau \in (t, t + \Delta t)$  and  $\xi \in (x, x + h)$ , so that  $T(x, t)$  in (9.18) has the expression

$$\begin{aligned}\frac{1}{\Delta t} T(x, t) &= -\frac{\Sigma(x + h) - \Sigma(x - h)}{2h} \frac{1}{2h} \left[ \frac{2h^3}{6} \frac{\partial^3 U(x, t)}{\partial x^3} + \frac{2h^5}{120} \frac{\partial^5 U(\xi, t)}{\partial x^5} \right] \\ &\quad + \frac{\Sigma(x)}{h^2} \left[ \frac{h^4}{12} \frac{\partial^4 U(x, t)}{\partial x^4} + \frac{h^6}{360} \frac{\partial^6 U(\xi, t)}{\partial x^6} \right] \\ &\quad + \frac{\Delta t}{2} \frac{\partial^2 U(x, t)}{\partial t^2} - \frac{\Delta t^2}{6} \frac{\partial^3 U(x, \tau)}{\partial t^3},\end{aligned}\quad (9.23)$$

which implies

$$|T(x, t)| \leq \Delta t [c_1(h) + c_2(\Delta t)], \quad (9.24)$$

where

$$\begin{aligned}c_1(h) &= (b - a) \left( \frac{hM_3}{12} + \frac{h^3M_5}{240} \right) + b \left( \frac{h^2M_4}{12} + \frac{h^4M_6}{360} \right) \\ c_2(\Delta t) &= \Delta t \left( \frac{N_2}{2} + \frac{\Delta t N_3}{6} \right),\end{aligned}\quad (9.25)$$

$|\partial^q U(x, t)/\partial x^q| \leq M_q$ ,  $q = 2, \dots, 6$ ,  $|\partial^r U(x, t)/\partial t^r| \leq N_r$ ,  $r = 2, 3$ , and  $M_q, N_r$  are positive constants. Since the right side of (9.24) does not depend on  $(x, t)$ , the constant  $M$  in (9.21) can be set equal to  $\Delta t [c_1(h) + c_2(\Delta t)]$ , so that  $\mathcal{E}(t + \Delta t) \leq \mathcal{E}(t) + \Delta t [c_1(h) + c_2(\Delta t)]$  or  $(\mathcal{E}(t + \Delta t) - \mathcal{E}(t))/\Delta t \leq c_1(h) + c_2(\Delta t)$ , which gives (9.17) since  $c_2(\Delta t) \rightarrow 0$  as  $\Delta t \rightarrow 0$ . Accordingly, we have  $\mathcal{E}(t) \leq c_1(h)t$ , that is, the error between the exact and finite difference solutions increases linearly in time for a given  $h$ . Since  $c_1(h) \rightarrow 0$  as  $h \rightarrow 0$  by (9.25),  $\mathcal{E}(t)$  can be made as small as desired in any bounded time interval by an appropriate selection of mesh size  $h$ .

Similar arguments can be used to develop bounds on the error  $\mathcal{E}(t)$  for transport equations defined on a bounded domain  $D \subset \mathbb{R}^d$ ,  $d \geq 2$ . As for the one-dimensional case, the resulting bound on  $\mathcal{E}(t)$  satisfies an inequality similar to that in (9.17).  $\blacktriangle$

The finite difference scheme in (9.16) yields the recurrence formula

$$Y(t + \Delta t) = (I + A \Delta t) Y(t) + W(t) \Delta t, \quad t = n \Delta t, \quad (9.26)$$

where  $Y(t)$  is a vector-valued process with coordinates  $Y_k(t) = U^*(x_k, t)$ ,  $A$  denotes a random matrix resulting from (9.16),  $I$  is the identity matrix, and  $W(t)$  is a vector with coordinates  $\{V(x_k, t)\}$ . The limit of (9.26) as  $\Delta t \rightarrow 0$  gives the ordinary differential equation

$$\dot{Y}(t) = A Y(t) + W(t), \quad t \geq 0, \quad (9.27)$$

with initial state  $Y_k(0) = U(x_k, 0) = U^*(x_k, 0)$ . Note that (9.27) can be obtained directly from (9.15) by approximating only the spatial derivatives in this equation by finite differences and that (9.26) results by integrating (9.27) over time intervals of length  $\Delta t$ . The vector  $Y(t)$  defined by (9.26) and (9.27) can be viewed as the state of discrete and continuous time linear systems with random coefficients of the type studied in Sects. 7.3 and 7.4.

The recurrence formula (9.16) and its matrix version (9.26) are useful only if their solutions can be used to approximate  $U(x, t)$  defined by (9.15), that is,  $\{U_k^n\}$  converges in some sense to  $U(x, t)$  as the spatial-temporal lattice used to construct the finite difference operator is refined. The classical Lax–Richtmyer theorem states that consistency and stability of a finite difference operator are necessary and sufficient conditions for the convergence of  $\{U_k^n\}$  to  $U(x, t)$  for well-posed, deterministic, linear partial differential equations ([10], Chap. 2). An operator  $\mathcal{L}_k^n$  is consistent if the truncation error  $\mathcal{L}[U(x_k, t_n)] - \mathcal{L}_k^n[U_k^n]$  converges to 0 as  $h \rightarrow 0$  and  $\Delta t \rightarrow 0$ . We say that  $\mathcal{L}_k^n$  is stable if the recurrence formula relating future values  $U(\cdot, t + \Delta t)$  to current values  $U(\cdot, t)$  generated by this operator does not amplify perturbations in data. The finite difference scheme in (9.16) has this property if, for example, a norm  $\|I + A \Delta t\|$  of matrix  $I + A \Delta t$  in (9.26) that is compatible with the norm of  $Y(t)$  satisfies the condition  $\|I + A \Delta t\| < 1$  or, equivalently, the eigenvalues of  $I + A \Delta t$  are included in the unit disc centered at the origin of the complex plane ([10], Chap. 2).

The proof of Theorem 9.1 is a direct application of arguments used to establish the convergence of finite difference schemes for deterministic, linear partial differential equations. We simply apply these arguments to almost all samples of the random conductivity field. The following example uses an extension of the classical Lax–Richtmyer theorem, referred to as the stochastic mean square Lax equivalence theorem [14], stating that a consistent approximating scheme for a SPDE is convergent in the mean square sense if and only if it is stable [14]. The conditions in the stochastic Lax theorem are weaker than those in Theorem 9.1 since they are for the mean square convergence of the sequence of finite difference solutions.

*Example 9.3* Let  $U(x, t)$  be the solution of partial differential equation

$$U(x, t) - U(x, 0) + a \int_0^t \frac{\partial U(x, s)}{\partial x} ds + b \int_0^t U(x, s) dB(s) = 0, \quad t \in [0, \tau], \quad x \in \mathbb{R}, \quad (9.28)$$

with initial state  $U(x, 0) = f(x)$ , where  $f(x)$  denotes a specified deterministic function and  $B$  is a Brownian motion. The finite difference approximation of this equation is

$$U_k^{n+1} = (1 + R) U_k^n - R U_{k+1}^n - b U_k^n \Delta B_n, \quad n = 0, 1, \dots, \quad k \in \mathbb{Z}, \quad (9.29)$$

where  $U_k^n$  is an approximation for  $U(k \Delta x, n \Delta t)$ ,  $\Delta x$  and  $\Delta t$  denote the space and time steps of the finite difference mesh,  $\partial U(k \Delta x, n \Delta t) / \partial t \simeq (U_k^{n+1} - U_k^n) / \Delta t$ ,  $\partial U(k \Delta x, n \Delta x) / \partial x \simeq (U_{k+1}^n - U_k^n) / \Delta x$ ,  $R = a \Delta t / \Delta x$ , and  $\Delta B_n = B((n+1) \Delta t) - B(n \Delta t)$ . It is assumed that  $a$ ,  $\Delta x$ , and  $\Delta t$  are such that  $R \in (-1, 0)$ .

Let  $\|\xi\|_{\ell^2, \Delta x}^2 = \sum_{k=-\infty}^{\infty} |\xi_k|^2 \Delta x$  denotes the  $\ell^2, \Delta x$ -norm of a countable sequence  $\xi = (\dots, \xi_{-1}, \xi_0, \xi_1, \dots)$  of reals. The finite difference scheme is stable in this norm since

$$E[\|U^{n+1}\|_{\ell^2, \Delta x}^2] \leq (1 + b^2 \Delta t) E[\|U^n\|_{\ell^2, \Delta x}^2] \leq (1 + b^2 \Delta t)^{n+1} E[\|U^0\|_{\ell^2, \Delta x}^2],$$

so that

$$E[\|U^{n+1}\|_{\ell^2, \Delta x}^2] \leq \left(1 + \frac{b^2 \Delta t}{n+1}\right)^{n+1} E[\|U^0\|_{\ell^2, \Delta x}^2] \leq e^{b^2 \Delta t} E[\|U^0\|_{\ell^2, \Delta x}^2]$$

for  $\Delta t = t/(n+1)$ .

Let  $\phi(x, t)$  be a smooth function, that is, a function that is continuously differentiable in  $x$  and continuous in  $t$ . The operator  $\mathcal{L}$  defined by (9.28) applied in a time interval  $(n \Delta t, (n+1) \Delta t)$  at node  $x_k = k \Delta x$  and the associated finite difference operator are

$$\begin{aligned} \mathcal{L}[\phi]_k^n &= \phi(x_k, (n+1) \Delta t) - \phi(x_k, n \Delta t) + a \int_{n \Delta t}^{(n+1) \Delta t} \frac{\partial \phi(x_k, s)}{\partial x} ds \\ &\quad + b \int_{n \Delta t}^{(n+1) \Delta t} \phi_x(x_k, s) dB(s) \quad \text{and} \\ \mathcal{L}_k^n[\phi] &= \phi(x_k, (n+1) \Delta t) - \phi(x_k, n \Delta t) + R (\phi_x(x_{k+1}, n \Delta t) - \phi_x(x_k, n \Delta t)) \\ &\quad + b \phi(x_k, n \Delta t) \Delta B_n. \end{aligned}$$

The expressions of  $\mathcal{L}$  and  $\mathcal{L}_k^n$  can be used to prove the convergence  $E[|\mathcal{L}[\phi]_k^n - \mathcal{L}_k^n[\phi]|^2] \rightarrow 0$  as  $\Delta x, \Delta t \rightarrow 0$ , which shows that the finite difference operator  $\mathcal{L}_k^n$  is consistent in the mean square sense. We summarize the essential arguments used to show that  $\mathcal{L}_k^n$  is consistent. The complete proof can be found in [15].  $\diamond$



*Proof* The recurrence formula in (9.29) with  $R \in (-1, 0)$  gives

$$\begin{aligned}
 E[|U_k^{n+1}|^2] &= E[(1+R)U_k^n - R U_{k+1}^n - b U_k^n \Delta B_n]^2 \\
 &= E[(1+R)U_k^n - R U_{k+1}^n]^2 + b^2 \Delta t E[|U_k^n|^2] \\
 &\leq (1+R)^2 E[|U_k^n|^2] + R^2 E[|U_{k+1}^n|^2] + 2|(1+R)R| E[|U_k^n U_{k+1}^n|] + b^2 \Delta t E[|U_k^n|^2] \\
 &\leq [(1+R)^2 + |(1+R)R| + b^2 \Delta t] E[|U_k^n|^2] + [R^2 + |(1+R)R|] E[|U_{k+1}^n|^2]
 \end{aligned} \tag{9.30}$$

by using  $2 E[|U_k^n U_{k+1}^n|] \leq E[|U_k^n|^2] + E[|U_{k+1}^n|^2]$ . Since  $|1+R| + |R| = 1$  and  $|R| < 1$ , we have

$$\sum_{k=-\infty}^{\infty} E[|U_k^{n+1}|^2] \leq (1+b^2 \Delta t) \sum_{k=-\infty}^{\infty} E[|U_k^n|^2], \tag{9.31}$$

so that  $E[\|U^{n+1}\|_{\ell_{2,\Delta x}}^2] \leq (1+b^2 \Delta t) E[\|U^n\|_{\ell_{2,\Delta x}}^2]$ .

The mean square discrepancy between the operators  $\mathcal{L}[\phi]_k^n$  and  $\mathcal{L}_k^n[\phi]$  is

$$\begin{aligned}
 E[|\mathcal{L}[\phi]_k^n - \mathcal{L}_k^n[\phi]|^2] &= E\left[\left|a \int_{n\Delta t}^{(n+1)\Delta t} \left(\phi_x(x_k, s) - \frac{\phi_{k+1}^n - \phi_k^n}{\Delta x}\right) ds \right. \right. \\
 &\quad \left. \left. + b \int_{n\Delta t}^{(n+1)\Delta t} (\phi(x_k, s) - \phi_k^n) dB(s) \right|^2\right] \\
 &\leq 2a^2 E\left[\left|\int_{n\Delta t}^{(n+1)\Delta t} \left(\phi_x(x_k, s) - \frac{\phi_{k+1}^n - \phi_k^n}{\Delta x}\right) ds\right|^2\right] \\
 &\quad + 2b^2 E\left[\left|\int_{n\Delta t}^{(n+1)\Delta t} (\phi(x_k, s) - \phi_k^n) dB(s)\right|^2\right],
 \end{aligned} \tag{9.32}$$

so that  $E[|\mathcal{L}[\phi]_k^n - \mathcal{L}_k^n[\phi]|^2] \rightarrow 0$  as  $\Delta x, \Delta t \rightarrow 0$  since the latter expectation in (9.32) is smaller than  $E[\int_{n\Delta t}^{(n+1)\Delta t} |\phi(x_k, s) - \phi_k^n|^2 ds]$ .

For the mean square convergence of the finite difference scheme, note that the difference  $Z_k^n = U(x_k, n\Delta t) - U_k^n$  satisfies the recurrence formula

$$\begin{aligned}
 Z_k^{n+1} &= Z_k^n - a \int_{n\Delta t}^{(n+1)\Delta t} \left(\frac{\partial U(x_k, s)}{\partial x} - \frac{U_{k+1}^n - U_k^n}{\Delta x}\right) ds \\
 &\quad - b \int_{n\Delta t}^{(n+1)\Delta t} (U(x_k, s) - U_k^n) dB(s),
 \end{aligned}$$

where

$$\begin{aligned}
 U(x_k, (n+1)\Delta t) &= U(x_k, n\Delta t) - a \int_{n\Delta t}^{(n+1)\Delta t} \frac{\partial U(x, s)}{\partial x} \Big|_{x=x_k} \\
 &\quad - b \int_{n\Delta t}^{(n+1)\Delta t} U(x_k, s) dB(s).
 \end{aligned}$$

Straightforward calculations and facts from [16, 17] are used in [15] to show that there exists a constant  $c > 0$  such that  $E[\|Z^{n+1}\|_{\ell_{2,\Delta x}}] \leq 4c(1 + (1+n\Delta t)e^{b^2 n \Delta t})\Delta t$ , so that  $E[\|Z^{n+1}\|_{\ell_{2,\Delta x}}] \rightarrow 0$  as  $\Delta t \rightarrow 0$ .  $\blacktriangle$

## 9.4 Applied SPDEs: Arbitrary Uncertainty

The SPDEs encountered in applications are usually more general than those considered in the mathematical literature. They are used to describe a broad range of physical phenomena, and their coefficients, input processes, and/or end conditions are Gaussian or non-Gaussian random functions of time and space, that are much more general than spatially dependent Gaussian white noise processes (Definition 9.2). Advance concepts of random functions and Itô's calculus are primarily used in the mathematical literature to solve SPDEs (Sect. 9.2). In the applied literature, the Monte Carlo simulation, stochastic Galerkin, and collocation methods are usually employed to solve SPDEs. A recent alternative to these methods is offered by stochastic reduced order models (SROMs). Taylor, perturbation, Neumann series, and related methods provide useful and simple solutions for SPDEs with random entries of small uncertainty (Sect. 9.5).

Let  $\mathcal{L}$  be a random differential operator, that is, a differential operator with random coefficients defined on a probability space  $(\Omega, \mathcal{F}, P)$  that involves both spatial and temporal derivatives, and let  $U(x, t)$  be the solution of the initial-boundary value problem

$$\mathcal{L}[U(x, t)] = V(x, t), \quad x \in D, \quad t \in (0, \tau), \quad (9.33)$$

with random coefficients and input, where the input  $V(x, t)$  is a random element on  $(\Omega, \mathcal{F}, P)$ ,  $D$  denotes a bounded open subset of  $\mathbb{R}^d$ , and  $(0, \tau)$  is a bounded time interval. The differential equation in (9.33) is accompanied by initial and boundary conditions that may be deterministic or random. If  $\mathcal{L}$  and  $V$  are time invariant and  $\mathcal{L}$  involves only spatial derivatives, then (9.33) becomes a stochastic boundary value problem denoted by  $\mathcal{L}[U(x)] = V(x)$ .

Analytical solutions for (9.33) and its time-invariant version  $\mathcal{L}[U(x)] = V(x)$  are rarely possible. Numerical methods need to be used for solution. Since these methods can only solve problems with finite numbers of degrees of freedom, it is only possible to solve approximate versions of (9.33), since the solution of (9.33) has an infinite number of degrees of freedom. Finite element, finite difference, and other methods can be employed to discretize the physical space. For example, (9.33) becomes an ordinary differential equation (ODE) with respect to time if its spatial derivatives are approximated by finite differences. The resulting ODE has random coefficients and input (Sect. 7.4). If the time argument is also discretized, the ODE becomes a recurrence formula with random coefficients and input (Sect. 7.3). The spatial discretization of the stochastic boundary value problem  $\mathcal{L}[U(x)] = V(x)$  results in a stochastic algebraic equation (SAE) of the type discussed in Chap. 8.

The solution of approximations of (9.33) obtained by discretization of the physical space by the methods in Chaps. 7 and 8 is likely to be inefficient since the number of random variables in the resulting approximate equations can be very large. For example, let  $\Sigma(x)$  be a random field in the definition of (9.33). The ODE derived from this equation by approximating all spatial derivatives by finite differences depends on the values of  $\Sigma(x)$  at the nodes of the finite difference scheme. The number of nodes can be excessive if high resolution meshes in space are required and/or the domain of definition of (9.33) is a subset of  $\mathbb{R}^3$  rather than an interval of the real line. Moreover, numerical difficulties may be encountered since values of  $\Sigma(x)$  at closely spaced nodes are likely to be strongly dependent. It is necessary to also discretize the probability space, which means to represent the random elements in the definition of (9.33) by parametric models that must (1) depend on a relatively small number of random variables for computational reasons and (2) satisfy both mathematical and physical constraints to guarantee the solution existence and uniqueness and provide realistic representations for physical properties (Sect. 6.3.3).

The solutions of SPDEs use different representations for the random elements in the definition of these equations. In the Monte Carlo method, the random elements in (9.33) are represented by a relatively large number of independent, equally likely samples of these elements. In the method based on stochastic reduced order models (SROMs), the random elements are represented by a relatively small number of samples that are not equally likely. In the stochastic Galerkin method, the random elements are approximated by parametric models usually obtained by truncating Karhunen–Loève representations of these elements and, subsequently, describing the random variables  $\{Z_i\}$  in these representations by truncated polynomial chaos expansions. In the stochastic collocation method, approximations are constructed for  $U(x, t)$  by interpolating between deterministic solutions of (9.33) corresponding to selected values of  $\{Z_i\}$  in the range of these variables.

### 9.4.1 General Considerations

Let  $U(x)$  be the solution of a stochastic boundary value problem  $\mathcal{L}[U(x)] = V(x)$ ,  $x \in D$ , accompanied by conditions that  $U$  must satisfy on the boundary  $\partial D$  of  $D$ , where  $D \subset \mathbb{R}^d$  is an open bounded subset of  $\mathbb{R}^d$ . It is assumed  $\mathcal{L}$  is a linear differential operator defined by

$$\mathcal{L}[U(x)] = \sum_{|\alpha|, |\beta| \leq m} (-1)^{|\alpha|} \mathcal{D}^\alpha (a_{\alpha\beta}(x) \mathcal{D}^\beta U(x)), \quad (9.34)$$

where  $a_{\alpha\beta} : D \rightarrow \mathbb{R}$  are real-valued functions,  $\alpha = (\alpha_1, \dots, \alpha_d)$ ,  $\alpha_i \geq 0$  are integers,  $|\alpha| = \alpha_1 + \dots + \alpha_d = 0, 1, \dots, m$ , and  $\mathcal{D}^\alpha = \partial^{|\alpha|} / \partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}$  denotes the  $\alpha$ -th weak derivative with respect to the spatial arguments of  $U$  ([18], Chap. 7). For a second order problem,  $m = 1$  so that  $\mathcal{L}$  has the form

$$\mathcal{L}[U(x)] = a_0(x) U(x) - \sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left( a_{ij}(x) \frac{\partial U(x)}{\partial x_j} \right) + \sum_{i=1}^d a_i(x) \frac{\partial U(x)}{\partial x_i} \quad (9.35)$$

Recall that a function  $f \in L^1_{\text{loc}}(D)$  has a weak derivative  $\mathcal{D}^\alpha f$  if there exists  $g \in L^1_{\text{loc}}(D)$  such that  $\int_D g(x) \phi(x) dx = (-1)^{|\alpha|} \int_D f(x) \phi^{(\alpha)}(x) dx$  holds for all  $\phi \in C_0^\infty(D)$ , where  $L^1_{\text{loc}}(D)$  is the set of functions  $f : D \rightarrow \mathbb{R}$  with the property  $f \in L^1(K)$  for all compacts in the interior of  $D$  and  $\phi \in C^\infty(D)$  with compact support ([18], Chap. 11).

**Definition 9.3** The operator  $\mathcal{L}$  in (9.34) is elliptic if  $\sum_{|\alpha|, |\beta|=m} a_{\alpha\beta}(x) \xi^{\alpha+\beta} \neq 0$  for all  $x \in D$  and arbitrary  $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$ ,  $\xi \neq 0$ , where  $\alpha_i \geq 0$ ,  $i = 1, \dots, d$ , are integers and  $\xi^\alpha$  is a short hand notation for  $\xi_1^{\alpha_1} \dots \xi_d^{\alpha_d}$ . The operator  $\mathcal{L}$  is said to be strongly elliptic if there exists  $\eta > 0$  such that  $|\sum_{|\alpha|, |\beta|=m} a_{\alpha\beta}(x) \xi^{\alpha+\beta}| \geq \eta \|\xi\|_{\mathbb{R}^d}^2$ .

Prior to considering partial differential equations with operators given by (9.34), we define a set of functions  $H^m(D)$  that constitutes a natural home for the solution of these equations. The set  $H^m(D)$ , referred to as Sobolev space, consists of real-valued functions defined on  $D$  with the properties

$$H^m(D) = \{U : D \rightarrow \mathbb{R}, \mathcal{D}^\alpha U \in L^2(D) \text{ for } |\alpha| \leq m\}, \quad (9.36)$$

where  $m \geq 0$  is an integer and  $L^2(D)$  denotes the set of real-valued square integrable functions defined on  $D$ . Note that  $H^n(D) \subset H^m(D)$  for  $n \geq m$  and that  $H^0(D) = L^2(D)$ .

**Theorem 9.2** *The set  $H^m(D)$  with the norm*

$$\|U\|_{H^m(D)}^2 = \langle U, U \rangle_{H^m(D)} = \sum_{|\alpha| \leq m} \|\mathcal{D}^\alpha U\|_{L^2(D)}^2 \quad (9.37)$$

*is a Hilbert space, where  $\|\cdot\|_{H^m(D)}$  is induced by the inner product*

$$\langle U, V \rangle_{H^m(D)} = \int_D \sum_{|\alpha| \leq m} (\mathcal{D}^\alpha U(x)) (\mathcal{D}^\alpha V(x)) dx = \sum_{|\alpha| \leq m} \langle \mathcal{D}^\alpha U, \mathcal{D}^\alpha V \rangle_{L^2(D)} \quad (9.38)$$

*and  $\langle \cdot, \cdot \rangle_{L^2(D)}$  denotes the inner product in  $L^2(D)$ .*

*Proof* Simple arguments and the definitions in (B.8) and (B.15) show that  $H^m(D)$  is a linear space and that  $\langle \cdot, \cdot \rangle_{H^m(D)}$  is an inner product on it. It remains to show that  $H^m(D)$  with the norm in (9.37) is complete.

Let  $\{U_p\}$  be a Cauchy sequence in  $H^m(D)$ , that is,  $\|U_p - U_q\|_{H^m(D)} \rightarrow 0$  as  $p, q \rightarrow \infty$ . This implies  $\|\mathcal{D}^\alpha(U_p - U_q)\|_{L^2(D)} \rightarrow 0$  as  $p, q \rightarrow \infty$  by the second equality in (9.37), so that  $\{\mathcal{D}^\alpha U_p\}$  is Cauchy in  $L^2(D)$  for each  $|\alpha| \leq m$ . Since  $L^2(D)$  is complete,  $\mathcal{D}^\alpha U_p$  converges to a function  $U^{(\alpha)} \in L^2(D)$ . We have

$$\begin{aligned}
\int_D U^{(\alpha)}(x) \phi(x) dx &= \int (\text{l.i.m.}_{p \rightarrow \infty} \mathcal{D}^\alpha U_p(x)) \phi(x) dx = \langle \text{l.i.m.}_{p \rightarrow \infty} \mathcal{D}^\alpha U_p, \phi \rangle_{L^2(D)} \\
&= \lim_{p \rightarrow \infty} \langle \mathcal{D}^\alpha U_p, \phi \rangle_{L^2(D)} = (-1)^{|\alpha|} \lim_{p \rightarrow \infty} \langle U_p, \mathcal{D}^\alpha \phi \rangle_{L^2(D)} \\
&= (-1)^{|\alpha|} \langle \text{l.i.m.}_{p \rightarrow \infty} U_p, \mathcal{D}^\alpha \phi \rangle = (-1)^{|\alpha|} \int_D U(x) \mathcal{D}^\alpha \phi(x) dx
\end{aligned} \tag{9.39}$$

by the continuity of the inner product and properties of generalized partial derivatives, where l.i.m. denotes mean square limit and  $\phi : D \rightarrow \mathbb{R}$  is an arbitrary infinitely differentiable function with compact support ([8], Sect. 1.2, [18], Chap. 7). The equality  $\int_D U^{(\alpha)}(x) \phi(x) dx = (-1)^{|\alpha|} \int_D U(x) \mathcal{D}^\alpha \phi(x) dx$  shows that  $U^{(\alpha)}$  is the weak derivative of order  $\alpha$  of  $U$  since this function and all its derivatives of order  $|\alpha| \leq m$  are in  $L^2(D)$ , so that  $U \in H^m(D)$ .  $\blacktriangle$

### 9.4.2 Deterministic Boundary Value Problems

Consider the deterministic boundary value problem

$$\mathcal{L}[U(x)] = V(x), \quad x \in D \subset \mathbb{R}^d, \tag{9.40}$$

with the boundary conditions (BCs)  $B_k U = g_k$ ,  $k = 0, 1, \dots, m-1$ , where  $\mathcal{L}$  is given by (9.34) and  $\{B_k\}$  denote linear differential operators. It is assumed that the boundary value problem (9.40) is well-posed, that is, it has a unique solution that depends continuously on data in the sense that there exists a constant  $c > 0$  such that  $\|U\|_{H^s(D)} \leq c \|V\|_{H^{s-2m}(D)}$  provided  $U \in H^s(D)$ . If (9.40) has this property and  $U_k(x)$  are solutions for data  $V_k(x)$ ,  $k = 1, 2$ , then  $\|U_1 - U_2\|_{H^s(D)} \leq c \|V_1 - V_2\|_{H^{s-2m}(D)}$  so that small changes in data can only yield small changes in the solution. A useful discussion on the existence, uniqueness, dependence on data, and regularity for solutions of elliptic boundary value problems can be found in ([18], Chap. 8). We are interested in conditions under which well-posed boundary value problems admit unique weak solutions.

In the remainder of this section, it is assumed that  $\mathcal{L}$  in (9.40) is an elliptic operator and that the essential boundary conditions for this equation are homogeneous, that is, boundary conditions involving differential operators of order strictly smaller than  $m$  ([18], Sect. 9.2). Let

$$\mathcal{V}^m(D) = \{W \in H^m(D) : W \text{ satisfies all essential BCs}\} \tag{9.41}$$

be the space of admissible functions. Arguments similar to those used to show that  $H^m(D)$  with the norm in (9.37) is a Hilbert space can be used to show that  $\mathcal{V}^m(D)$  with the same norm is a Hilbert space. The weak form of (9.40) is

$$\mathcal{A}(U, W) = \langle V, W \rangle, \quad \forall W \in \mathcal{V}^m(D), \tag{9.42}$$

where

$$\mathcal{A}(U, W) = \int_D \sum_{|\alpha|, |\beta| \leq m} a_{\alpha\beta}(x) \mathcal{D}^\alpha U(x) \mathcal{D}^\beta W(x) dx + \text{boundary terms}$$

is a bilinear form and  $\langle V, \cdot \rangle : \mathcal{V}^m(D) \rightarrow \mathbb{R}$  is a linear functional ([18], Sect. 9.2). This form results by multiplying both sides of (9.40) with  $W \in \mathcal{V}^m(D)$ , integrating over  $D$ , and using Green's theorem.

Our objective is to find  $U \in \mathcal{V}^m(D)$  satisfying (9.42), that is, the weak solution of (9.40). If the bilinear form  $\mathcal{A}(\cdot, \cdot)$  is bounded and elliptic, that is, if there exists constants  $M > 0$  and  $\beta > 0$  such that  $|\mathcal{A}(U, W)| \leq M \|U\|_{\mathcal{V}^m(D)} \|W\|_{\mathcal{V}^m(D)}$  and  $\mathcal{A}(W, W) \geq \beta \|W\|_{\mathcal{V}^m(D)}^2$  for all  $U, W \in \mathcal{V}^m(D)$ , then (9.42) admits a unique solution by the Lax–Milgram theorem (B.44). Moreover, the solution depends continuously on data ([18], Theorem 1, p. 316).

*Example 9.4* Let  $U : D \rightarrow \mathbb{R}$  be the solution of the boundary value problem

$$-\nabla \cdot (a(x) \nabla U(x)) = f(x), \quad x \in D. \quad (9.43)$$

with  $U(x) = 0$  for  $x \in \partial D$ . The weak form of this equation is

$$\mathcal{B}(U, W) = \langle f, W \rangle_{L^2(D)}, \quad \forall W \in \mathcal{W}(D), \quad (9.44)$$

where  $\langle f, W \rangle_{L^2(D)} = \int_D f(x) W(x) dx$ ,

$$\mathcal{W}(D) = \{W : D \rightarrow \mathbb{R}, W \in L^2(D), \nabla W \in L^2(D), \text{ and } W = 0 \text{ on } \partial D\} \quad (9.45)$$

is the linear space  $\mathcal{V}^1(D)$  in (9.41), and

$$\mathcal{B}(U, W) = \int_D a(x) \nabla U(x) \cdot \nabla W(x) dx, \quad U, W \in \mathcal{W}(D), \quad (9.46)$$

is a real-valued bilinear form defined on  $\mathcal{W}(D) \times \mathcal{W}(D)$ . Note that the left and the right sides of (9.44) correspond to  $\mathcal{A}(U, W)$  and  $\langle V, W \rangle$  given by (9.42) and that the inner product and norm on  $\mathcal{W}(D)$  are those on  $\mathcal{V}^1(D)$  in (9.41), that is,  $\langle U, W \rangle_{\mathcal{W}(D)} = \int_D (U W + \nabla U \cdot \nabla W) dx$  and  $\|U\|_{\mathcal{W}(D)}^2 = \int_D (U^2 + \nabla U \cdot \nabla U) dx$ . If  $f \in L^2(D)$  and  $a(x) \in [\alpha, \beta]$ ,  $0 < \alpha \leq \beta < \infty$ , for all  $x \in D$ , then (9.44) admits a unique solution.  $\diamond$

*Proof* The right side of (9.43) multiplied with  $W \in \mathcal{W}(D)$  and integrated over  $D$  gives  $\langle f, W \rangle = \int_D f(x) W(x) dx$ . The left side of (9.43) multiplied with  $W \in \mathcal{W}(D)$  and integrated over  $D$  gives

$$-\sum_{i=1}^d \int_D \frac{\partial}{\partial x_i} \left( a(x) \frac{\partial U(x)}{\partial x_i} \right) W(x) dx = \int_D a(x) \nabla U(x) \cdot \nabla W(x) dx, \quad \forall W \in \mathcal{W}(D),$$

by Green's theorem ([18], Sects. 7.2 and 7.4) that can be given in the form

$$\int_D u \frac{\partial v}{\partial x_i} dx = \int_{\partial D} u v n_i ds - \int_D v \frac{\partial u}{\partial x_i} dx, \quad (9.47)$$

where  $D$  is an open bounded set in  $\mathbb{R}^d$ ,  $n(x) = (n_1(x), \dots, n_d(x))$  denotes the outward unit normal to  $D$  at  $x \in \partial D$ , and  $u, v \in H^1(\overline{D})$ .

That  $\mathscr{W}(D)$  in (9.45) is a linear space and

$$\langle U, W \rangle_{\mathscr{W}(D)} = \int_D (U W + \nabla U \cdot \nabla W) dx$$

defines an inner product on this space follows by straightforward arguments. Since  $f \in L^2(D)$ , the right side of (9.44) is linear in  $W$  and bounded. The bilinear form  $\mathscr{B}$  is continuous since

$$\begin{aligned} |\mathscr{B}(U, W)| &= \left| \int_D (a(x)^{1/2} \nabla U(x)) \cdot (a(x)^{1/2} \nabla W(x)) dx \right| \\ &\leq \left( \int_D a(x) \nabla U(x) \cdot \nabla U(x) dx \int_D a(x) \nabla W(x) \cdot \nabla W(x) dx \right)^{1/2} \\ &\leq \beta \|U\|_{\mathscr{W}(D)} \|W\|_{\mathscr{W}(D)}, \quad \forall U, W \in \mathscr{W}(D), \end{aligned}$$

by the Cauchy–Schwarz inequality, the assumption  $a(x) \in [\alpha, \beta]$ ,  $0 < \alpha \leq \beta < \infty$ ,  $x \in D$ , and the inequality

$$\int_D \nabla W(x) \cdot \nabla W(x) dx \leq \int_D (W(x)^2 + \nabla W(x) \cdot \nabla W(x)) dx = \|W\|_{\mathscr{W}(D)}^2$$

for all  $W \in \mathscr{W}(D)$ .

The bilinear form in (9.46) is also elliptic since

$$\mathscr{B}(W, W) = \int_D a(x) \nabla W(x) \cdot \nabla W(x) dx \geq \alpha \int_D \nabla W(x) \cdot \nabla W(x) dx = \alpha \|\nabla W\|_{L^2(D)}^2$$

by properties of  $a(x)$ . Also, there exists a constant  $c > 0$  such that  $\int_D W^2 dx \leq c \int_D \nabla W \cdot \nabla W dx$  by the Poincaré–Friedrichs inequality ([18], Theorem 9, Chap. 7), so that  $\|W\|_{\mathscr{W}(D)}^2 = \|W\|_{L^2(D)}^2 + \|\nabla W\|_{L^2(D)}^2 \leq (1 + c) \|\nabla W\|_{L^2(D)}^2$ , which gives  $\|\nabla W\|_{L^2(D)}^2 \geq \|W\|_{\mathscr{W}(D)}^2 / (1 + c)$ . The latter inequality and  $\mathscr{B}(W, W) \geq \alpha \|\nabla W\|_{L^2(D)}^2$  imply  $\mathscr{B}(W, W) \geq [\alpha / (1 + c)] \|W\|_{\mathscr{W}(D)}^2$ . Then (9.44) has a unique solution by the Lax–Milgram theorem (Theorem B.44).  $\blacktriangle$

### 9.4.3 Stochastic Boundary Value Problems

Consider now the stochastic version of (9.40), that is, the coefficients of the differential operator  $\mathscr{L}$ , the right side, and/or the boundary conditions of this equation depend on random elements defined on a probability space  $(\Omega, \mathscr{F}, P)$ , so that the resulting solution  $U : D \times \Omega \rightarrow \mathbb{R}$  is a function of  $x \in D$  and  $\omega \in \Omega$ . Let  $(D \times \Omega, \mathscr{B}(D) \times \mathscr{F}, \lambda \times P)$  denote the product of measure spaces  $(D, \mathscr{B}(D), \lambda)$  and  $(\Omega, \mathscr{F}, P)$ , where  $\mathscr{B}(D)$  is the Borel  $\sigma$ -field on  $D$  and  $\lambda$  denotes the Lebesgue measure on  $\mathbb{R}^d$ . The analog of the space of admissible functions  $\mathscr{V}^m(D)$  in (9.41) for stochastic boundary value problems is the set

$$\mathcal{V}^m(D, \Omega) = \{W : D \times \Omega \rightarrow \mathbb{R}, \mathcal{D}^\alpha U \in L^2(D \times \Omega, \mathcal{B}(D) \times \mathcal{F}, \lambda \times P), |\alpha| \leq m, \\ \mathcal{B}(D) \times \mathcal{F}\text{-measurable, and satisfies } P\text{-a.s. all essential BCs}\}, \quad (9.48)$$

where  $\mathcal{D}^\alpha U$  denotes the  $\alpha$ -th weak derivative of  $U$  with respect to spatial arguments ([18], Chap. 7) and  $\lambda(dx) = dx$  is the infinitesimal volume in the physical space.

**Theorem 9.3** *The set of functions  $\mathcal{V}^m(D, \Omega)$  with the norm*

$$\|U\|_{\mathcal{V}^m(D, \Omega)}^2 = E \left[ \sum_{|\alpha| \leq m} \|\mathcal{D}^\alpha U\|_{L^2(D)}^2 \right] \quad (9.49)$$

is a Hilbert space, where  $\|\cdot\|_{\mathcal{V}^m(D, \Omega)}$  is the norm induced by the inner product

$$\begin{aligned} \langle U, W \rangle_{\mathcal{V}^m(D, \Omega)} &= E \left[ \int_D \sum_{|\alpha| \leq m} (\mathcal{D}^\alpha U(x, \cdot)) (\mathcal{D}^\alpha W(x, \cdot)) dx \right] \\ &= E \left[ \sum_{|\alpha| \leq m} \langle \mathcal{D}^\alpha U, \mathcal{D}^\alpha W \rangle_{L^2(D)} \right] = \sum_{|\alpha| \leq m} \langle \mathcal{D}^\alpha U, \mathcal{D}^\alpha W \rangle_{L^2(D \times \Omega)}, \end{aligned} \quad (9.50)$$

and  $\langle \cdot, \cdot \rangle_{L^2(D \times \Omega)}$  denotes the inner product on  $L^2(D \times \Omega)$ .

*Proof* That  $\mathcal{V}^m(D, \Omega)$  is a linear space and (9.50) is an inner product on this space follows by checking that the defining properties of linear spaces and inner products hold. Since  $\mathcal{V}^m(D, \Omega)$  consists of  $\mathcal{B}(D) \times \mathcal{F}$ -measurable and  $\lambda \times P$ -integrable functions,  $U(x, \cdot)$  is  $\mathcal{F}$ -measurable and  $P$ -integrable,  $U(\cdot, \omega)$  is  $\mathcal{B}(D)$ -measurable and  $\lambda$ -integrable, the expectation  $\int_\Omega U(\cdot, \omega) P(d\omega)$  is  $\mathcal{B}(D)$ -measurable and  $\lambda$ -integrable, the integral  $\int_D U(x, \cdot) \lambda(dx)$  is  $\mathcal{F}$ -measurable and  $P$ -integrable, and the order of the integrals over  $D$  and  $\Omega$  can be interchanged by Fubini's theorem (Theorem 2.9).

Consider a Cauchy sequence  $\{U_p\}$  in  $\mathcal{V}^m(D, \Omega)$ , that is,  $\|U_p - U_q\|_{\mathcal{V}^m(D, \Omega)} \rightarrow 0$  as  $p, q \rightarrow \infty$ , which implies  $\|\mathcal{D}^\alpha(U_p - U_q)\|_{L^2(D \times \Omega)} \rightarrow 0$  as  $p, q \rightarrow \infty$  for  $|\alpha| \leq m$ . Since  $L^2(D \times \Omega)$  is complete,  $\mathcal{D}^\alpha U_p$  converges to an element  $U^{(\alpha)} \in L^2(D \times \Omega)$  as  $p \rightarrow \infty$ . Let  $\phi : D \times \Omega \rightarrow \mathbb{R}$  be as in (9.39) for almost all  $\omega \in \Omega$ . Calculations as in this equation give

$$E \left[ \int_D U^{(\alpha)}(x, \omega) \phi(x, \omega) \lambda(dx) \right] = E \left[ (-1)^{|\alpha|} \int_D U(x, \omega) \mathcal{D}^\alpha \phi(x, \omega) \lambda(dx) \right] \quad (9.51)$$

for  $|\alpha| \leq m$ , so that

$$\int_D U^{(\alpha)}(x, \omega) \phi(x, \omega) \lambda(dx) = (-1)^{|\alpha|} \int_D U(x, \omega) \mathcal{D}^\alpha \phi(x, \omega) \lambda(dx) \quad (9.52)$$

holds  $P$ -a.s., which shows that  $U^{(\alpha)}$  is the weak derivative of order  $\alpha$  of  $U$  for almost all  $\omega \in \Omega$  with respect to the spatial argument  $x \in D$ . Since  $U^{(\alpha)}$  is in  $L^2(D \times \Omega)$  for all  $|\alpha| \leq m$ , we have  $U \in \mathcal{V}^m(D, \Omega)$ .  $\blacktriangle$



*Example 9.5* Suppose the coefficients and the right side of (9.43) are random functions on a probability space  $(\Omega, \mathcal{F}, P)$ , so that its solution  $U$  is a random function defined on the same probability space that satisfies the stochastic boundary value problem

$$-\nabla \cdot (a(x, \omega) \nabla U(x, \omega)) = f(x, \omega), \quad (x, \omega) \in D \times \Omega, \quad (9.53)$$

with  $U(x, \omega) = 0$  for  $x \in \partial D$  and almost all  $\omega \in \Omega$  and  $f \in L^2(D \times \Omega)$ . The weak form of (9.53) is

$$\mathcal{B}(U, W) = \langle f, W \rangle_{L^2(D \times \Omega)}, \quad \forall W \in \mathcal{W}(D, \Omega), \quad (9.54)$$

where  $\langle f, W \rangle_{L^2(D \times \Omega)} = E \left[ \int_D f W dx \right]$ ,

$$\begin{aligned} \mathcal{W}(D, \Omega) = \{ & W : D \times \Omega \rightarrow \mathbb{R}, W \in L^2(D \times \Omega), \nabla W \in L^2(D \times \Omega), \\ & W = \mathcal{B}(D) \times \mathcal{F}\text{-measurable, and } W(x, \cdot) = 0, x \in \partial D, P\text{-a.s.} \} \end{aligned} \quad (9.55)$$

is a linear space, and

$$\mathcal{B}(U, W) = E \left[ \int_D a(x, \cdot) \nabla U(x, \cdot) \cdot \nabla W(x, \cdot) dx \right] \quad (9.56)$$

is a real-valued bilinear form defined on  $\mathcal{W}(D, \Omega) \times \mathcal{W}(D, \Omega)$ . The inner product and the norm on  $\mathcal{W}(D, \Omega)$  are those on  $\mathcal{V}^1(D, \Omega)$ , that is,  $\langle U, W \rangle_{\mathcal{W}(D, \Omega)} = E \left[ \int_D (U W + \nabla U \cdot \nabla W) dx \right]$  and  $\|W\|_{\mathcal{W}(D, \Omega)}^2 = E \left[ \int_D (W^2 + \nabla W \cdot \nabla W) dx \right]$ .

If the functions  $a, f : D \times \Omega \rightarrow \mathbb{R}$  are  $\mathcal{B}(D) \times \mathcal{F}$ -measurable and  $\lambda \times P$ -integrable,  $a(x, \omega) \in [\alpha, \beta]$ ,  $0 < \alpha \leq \beta < \infty$ , for all  $x \in D$  and almost all  $\omega \in \Omega$ , and  $f \in L^2(D \times \Omega)$ , then (9.54) admits a unique solution. Moreover, the solution of (9.54) is stable. Note that  $\mathcal{B}(D)$  and  $\mathcal{B}(U, W)$  in (9.56) have different meanings.  $\diamond$

*Proof* The functional  $\langle f, W \rangle_{\mathcal{W}(D, \Omega)} = E \left[ \int_D f(x) W(x) dx \right]$  of  $W \in \mathcal{W}(D, \Omega)$  is linear and bounded. The bilinear form in (9.56) is bounded since

$$\begin{aligned} |\mathcal{B}(U, W)| &= \left| \int_{D \times \Omega} a(x, \omega) \nabla U(x, \omega) \cdot \nabla W(x, \omega) \lambda(dx) P(d\omega) \right| \\ &\leq \left( E \left[ \int_D a \nabla U \cdot \nabla U dx \right] E \left[ \int_D a \nabla W \cdot \nabla W dx \right] \right)^{1/2} \\ &\leq \beta \|U\|_{\mathcal{W}(D, \Omega)} \|W\|_{\mathcal{W}(D, \Omega)}, \quad \forall U, W \in \mathcal{W}(D, \Omega) \end{aligned}$$

by the Cauchy–Schwarz inequality, the relationship  $\|\nabla W\|_{L^2(D \times \Omega)} \leq \|W\|_{\mathcal{W}(D, \Omega)}$ , and properties of  $a(x, \omega)$ . Arguments as in Example 9.4 show that this form also has the property  $\mathcal{B}(U, U) \geq c \|U\|_{\mathcal{W}(D, \Omega)}^2$ ,  $\forall U \in \mathcal{W}(D, \Omega)$ , where  $c > 0$  is a constant. The Lax–Milgram theorem (Theorem B.44) implies that (9.54) admits a unique solution (see also Sect. 9.4.8.3).

The solution stability follows from the inequality  $\mathcal{B}(U, U) \geq \alpha \|U\|_{\mathcal{W}(D, \Omega)}^2$  and the fact that  $\langle f, U \rangle$  is a linear continuous functional. The latter observation implies that  $\langle f, U \rangle$  is bounded (Theorem B.23), that is, there exists a constant  $c'' > 0$  such that  $|\langle f, U \rangle| \leq c'' \|U\|_{\mathcal{W}(D, \Omega)}$ ,  $U \in \mathcal{W}(D, \Omega)$ . These inequalities and (9.54) give  $c' \|U\|_{\mathcal{W}(D, \Omega)}^2 \leq \mathcal{B}(U, U) = |\mathcal{B}(U, U)| = |\langle f, W \rangle| \leq c'' \|U\|_{\mathcal{W}(D, \Omega)}$ , so that  $\|U\|_{\mathcal{W}(D, \Omega)} \leq c''/c'$ . Solution uniqueness follows from this condition since, if  $\{U_k\}$  are solutions, that is,  $\mathcal{B}(U_k, W) = \langle f, W \rangle_{L^2(D \times \Omega)}$ ,  $k = 1, 2$ , we have  $\mathcal{B}(U_1 - U_2, W) = 0$  for all  $W \in \mathcal{W}(D, \Omega)$  so that  $\|U_1 - U_2\|_{\mathcal{W}(D, \Omega)} = 0$ .  $\blacktriangle$

There is no conceptual difference between deterministic and stochastic boundary value problems. The Lax–Milgram theorem is used to establish conditions under which these problems admit unique weak solutions. The analogy between deterministic and stochastic problems extends to numerical solutions. For deterministic BVPs, the physical space needs to be discretized by, for example, finite differences or finite elements, such that the original problem with an uncountable number of degrees of freedom is approximated by a problem with a finite number of degrees of freedom. For stochastic BVPs, both the physical space and the probability space need to be discretized. The physical space is discretized as for deterministic problems. The discretization of the probability space is usually achieved by representing the random fields in the definition of these equations by parametric models, that is, deterministic functions of space depending on a finite number of random variables. Truncated Karhunen–Loève series are typically used to construct parametric models.

The essential differences between deterministic and stochastic boundary value problems relate to the spaces in which we seek solutions and the measures and norms employed in these spaces. The solution for deterministic BVPs belongs to a subspace of  $H^m(D)$  endowed with the Lebesgue measure and the norm in (9.37). The solution for stochastic BVPs belongs to a subspace of the product of two measure spaces, the physical space  $(D, \mathcal{B}(D), \lambda)$  and the probability space  $(\Omega, \mathcal{F}, P)$ . The measure on this space is the product measure  $\lambda \times P$ , rather than the Lebesgue measure, and the norm is given by (9.49).

We conclude this section with a brief discussion on the discrepancy between weak solutions of stochastic boundary value problems that have the same functional form but different coefficients and right sides. Let  $U$  be the solution of (9.54) and let  $\tilde{U}$  be the solution of  $\tilde{\mathcal{B}}(\tilde{U}, W) = E[\int_D \tilde{f} W dx]$ , where  $\tilde{\mathcal{B}}$  is defined by (9.56) with  $\tilde{a}$  in place of  $a$ . It is assumed that  $\tilde{a}$  and  $\tilde{f}$  have the same properties as  $a$  and  $f$  and are defined on the same probability space as these random functions, so that  $\tilde{\mathcal{B}}(\tilde{U}, W) = E[\int_D \tilde{f} W dx]$  admits a unique solution. The following result relates differences between  $U$  and  $\tilde{U}$  to those between  $(a, f)$  and  $(\tilde{a}, \tilde{f})$ .

**Theorem 9.4** *The discrepancy between  $U$  and  $\tilde{U}$  can be bounded by*

$$\|U - \tilde{U}\|_{\mathcal{W}(D, \Omega)} \leq \frac{1+c}{\alpha} \left( \|a - \tilde{a}\|_{L^\infty(D \times \Omega)} \|\tilde{U}\|_{\mathcal{W}(D, \Omega)} + \|f - \tilde{f}\|_{L^2(D \times \Omega)} \right), \quad (9.57)$$

where  $\alpha > 0$  and  $c > 0$  are constants such that  $P(\inf_{x \in D} \{a(x)\} > \alpha) = 1$ .

*Proof* Since  $U - \tilde{U} \in \mathcal{W}(D, \Omega)$ ,  $\mathcal{B}(U - \tilde{U}, W)$  is defined and

$$\begin{aligned}
|\mathcal{B}(U - \tilde{U}, W)| &= |\mathcal{B}(U, W) - \tilde{\mathcal{B}}(\tilde{U}, W) + \tilde{\mathcal{B}}(\tilde{U}, W) - \mathcal{B}(\tilde{U}, W)| \\
&\leq |\tilde{\mathcal{B}}(\tilde{U}, W) - \mathcal{B}(\tilde{U}, W)| + |\langle f - \tilde{f}, W \rangle_{L^2(D \times \Omega)}| \\
&\leq \left| E \left[ \int_D (\tilde{a} - a) \nabla \tilde{U} \cdot \nabla W \, dx \right] \right| + \|f - \tilde{f}\|_{L^2(D \times \Omega)} \|W\|_{L^2(D \times \Omega)} \\
&\leq \left( E \left[ \int_D (\tilde{a} - a)^2 \nabla \tilde{U} \cdot \nabla \tilde{U} \, dx \right] E \left[ \int_D \nabla W \cdot \nabla W \, dx \right] \right)^{1/2} \\
&\quad + \|f - \tilde{f}\|_{L^2(D \times \Omega)} \|W\|_{L^2(D \times \Omega)} \\
&\leq \|a - \tilde{a}\|_{L^\infty(D \times \Omega)} \|\tilde{U}\|_{\mathcal{W}(D, \Omega)} \|W\|_{\mathcal{W}(D, \Omega)} + \|f - \tilde{f}\|_{L^2(D \times \Omega)} \|W\|_{\mathcal{W}(D, \Omega)},
\end{aligned}$$

where we use  $\|\nabla V\|_{L^2(D \times \Omega)}^2 \leq \|V\|_{\mathcal{W}(D, \Omega)}^2$  and  $\|V\|_{L^2(D \times \Omega)}^2 \leq \|V\|_{\mathcal{W}(D, \Omega)}^2$ . The Poincaré–Friedrichs inequality ([18], Sect. 7.5, Theorem 9) implies the existence of a constant  $c > 0$  such that  $\|W\|_{L^2(D \times \Omega)}^2 \leq c \|\nabla W\|_{L^2(D \times \Omega)}^2$ , so that  $\|W\|_{\mathcal{W}(D, \Omega)}^2 = \|W\|_{L^2(D \times \Omega)}^2 + \|\nabla W\|_{L^2(D \times \Omega)}^2 \leq (1 + c) \|\nabla W\|_{L^2(D \times \Omega)}^2$ . Since  $\mathcal{B}(W, W) = E \left[ \int_D a \nabla W \cdot \nabla W \, dx \right] \geq \alpha \|\nabla W\|_{L^2(D \times \Omega)}^2$ , we have  $\mathcal{B}(W, W) \geq c' \|W\|_{\mathcal{W}(D, \Omega)}^2$ , where  $c' = \alpha/(1 + c)$ .

$$\begin{aligned}
c' \|U - \tilde{U}\|_{\mathcal{W}(D, \Omega)}^2 &\leq \mathcal{B}(U - \tilde{U}, U - \tilde{U}) \leq (\|a - \tilde{a}\|_{L^\infty(D \times \Omega)} \|\tilde{U}\|_{\mathcal{W}(D, \Omega)} \\
&\quad + \|f - \tilde{f}\|_{L^2(D \times \Omega)}) \|U - \tilde{U}\|_{\mathcal{W}(D, \Omega)},
\end{aligned}$$

which gives the bound in (9.57) following division by  $\|U - \tilde{U}\|_{\mathcal{W}(D, \Omega)}$ .  $\blacktriangle$

Consistent with our intuition, the discrepancy between  $U$  and  $\tilde{U}$  depends on differences between  $(a, f)$  and  $(\tilde{a}, \tilde{f})$ , and can be bounded by the sum of the norms  $\|a - \tilde{a}\|_{L^\infty(D \times \Omega)}$  and  $\|f - \tilde{f}\|_{L^2(D \times \Omega)}$  weighted with some constants. We can view  $\tilde{U}$  as an approximation of  $U$  that can be, for example, a numerical solution of (9.54).

Consider first the deterministic version (9.44) of (9.54), and suppose that the domain  $D$  is partition in triangular finite elements  $\{T_r\}$  with the largest mesh  $h = \max_r \{\text{diam}(T_r)\} > 0$ . The finite element space,

$$\mathcal{W}_h(D) = \{W : D \rightarrow \mathbb{R} \text{ continuous and linear in each } T_r \text{ and } W = 0 \text{ on } \partial D\}, \quad (9.58)$$

is a subspace of  $\mathcal{W}(D)$  so that  $\mathcal{B}(U_h, W) = \langle f, W \rangle_{L^2(D)}$ ,  $W \in \mathcal{W}_h(D)$ , has a unique solution. The discrepancy between the solutions  $U$  of (9.44) and its finite element approximation  $U_h \in \mathcal{W}_h(D)$  can be bounded by

$$\|U - U_h\|_{\mathcal{W}(D)} \leq c h, \quad (9.59)$$

where  $c > 0$  is a constant that does not depend on  $h$  ([9], Chap. 2). Arguments yielding this inequality extend directly to finite element solutions for the stochastic problem (9.54) provided the random fields  $a(x, \omega)$  and  $f(x, \omega)$  are parametric models depending on a random vector  $Z$  with finite dimension. These fields can be viewed as deterministic functions  $a(x, y)$  and  $f(x, y)$  of  $(x, y) \in D \times \Gamma$ ,  $\Gamma = Z(\Omega)$ , so that (9.54) can be interpreted as a deterministic equation that can be solved by standard

finite element algorithms using an extension of  $\mathcal{W}_h(D)$  that includes approximating functions defined on  $D \times \Gamma$  [20]. Numerical solutions under this formulation are feasible only for problems with small stochastic dimensions, that is, problems depending on vectors  $Z$  that have just a few coordinates [21].

#### 9.4.4 Monte Carlo Simulation

Samples  $U(x, t, \omega)$  of the solution  $U(x, t)$  of (9.33), that is, deterministic solutions of this equation corresponding to samples of  $\mathcal{L}$  and  $V$ , and end conditions, are used to estimate properties of  $U(x, t)$ . The generation of samples of the random elements in (9.33) are usually based on parametric models, that is, deterministic functions of space and/or time depending on a finite number of random variables that may or may not be independent. Methods for generating samples of random elements are in Sects. 2.13 and 3.8.

Let  $U^{(n)}(x, t)$  denote the solution of (9.33) with all its random elements approximated by parametric models depending on  $n$  random variables. Properties of  $U(x, t)$  are estimated from samples of  $U^{(n)}(x, t)$ . A Monte Carlo simulation algorithm is useful if  $U^{(n)}(x, t)$  converges in some sense to  $U(x, t)$  as the discretization of the probability space is refined, that is,  $n \rightarrow \infty$ . The performance of Monte Carlo estimators for properties of  $U(x, t)$  depends on the accuracy of the parametric models and the sample size, that is, the number of samples of  $U^{(n)}(x, t)$  used for estimation.

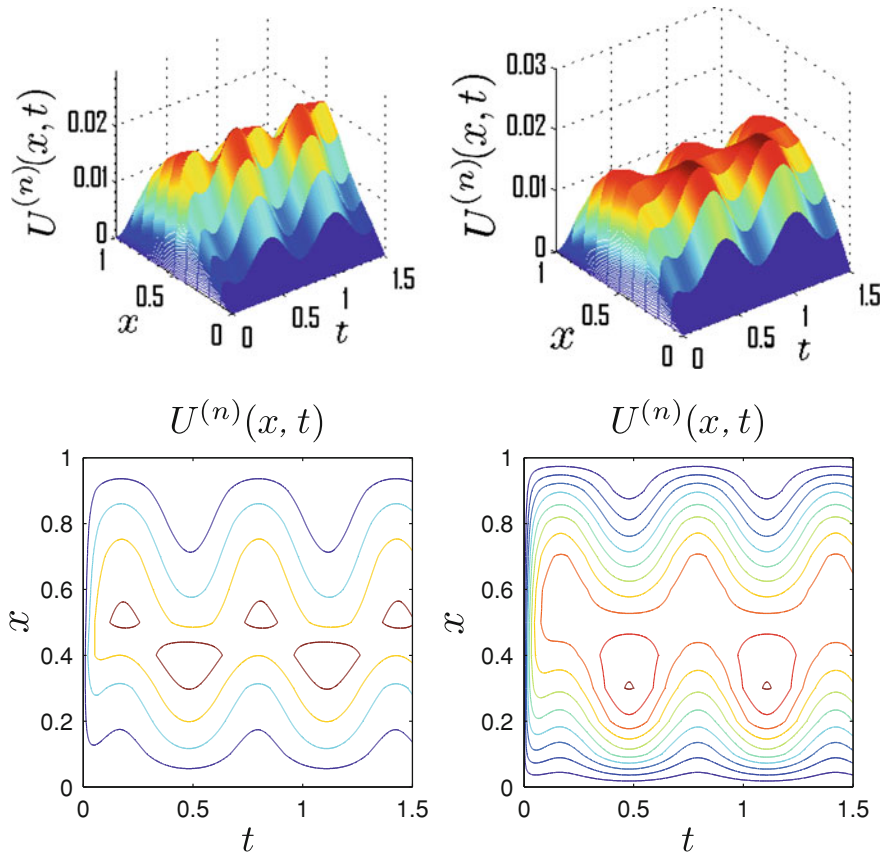
*Example 9.6* Let  $U(x, t)$  be a real-valued random function satisfying the stochastic partial differential equation (SPDE)

$$\frac{\partial U(x, t)}{\partial t} = \frac{\partial}{\partial x} \left( \Sigma(x) \frac{\partial U(x, t)}{\partial x} \right) + V(x, t), \quad x \in (0, l), \quad t \geq 0, \quad (9.60)$$

with boundary and initial conditions  $U(0, t) = U(l, t) = 0$ ,  $t \geq 0$ , and  $U(x, 0) = 0$ ,  $x \in (0, l)$ , where  $V(x, t) = \sin(vt)$ ,  $v > 0$ ,  $\Sigma(x) = F^{-1} \circ \Phi(G(x))$  is a translation random field,  $F$  denotes a distribution, and  $G(x)$  is a homogeneous Gaussian field with mean 0 and covariance function  $\rho(\xi) = E[G(x + \xi)G(x)] = (1 + \lambda |\xi|) \exp(-\lambda |\xi|)$ ,  $\lambda > 0$ , and one-sided spectral density  $g(v) = 4\lambda^3 / [\pi (v^2 + \lambda^2)^2]$ ,  $v \geq 0$ . This SPDE is a special case of (9.15) for  $d = 1$ .

The construction of a Monte Carlo algorithm for estimating properties of  $U(x, t)$  involves three steps. First, a sequence of homogeneous Gaussian fields  $G^{(n)}(x)$ ,  $n = 1, 2, \dots$ , with mean 0 and variance 1, that converges in some sense to  $G(x)$  is constructed and used to generate approximately samples of  $G(x)$ . Second, samples of  $G^{(n)}(x)$  are mapped into samples of  $\Sigma^{(n)}(x) = F^{-1} \circ \Phi(G^{(n)}(x))$ , that is, samples of a sequence of translation random fields approximating  $\Sigma(x)$ . Third, samples of the solution  $U^{(n)}(x, t)$  of (9.60) with  $\Sigma^{(n)}(x)$  in place of  $\Sigma(x)$  are calculated by deterministic solvers, and used subsequently to estimate properties of  $U(x, t)$ .

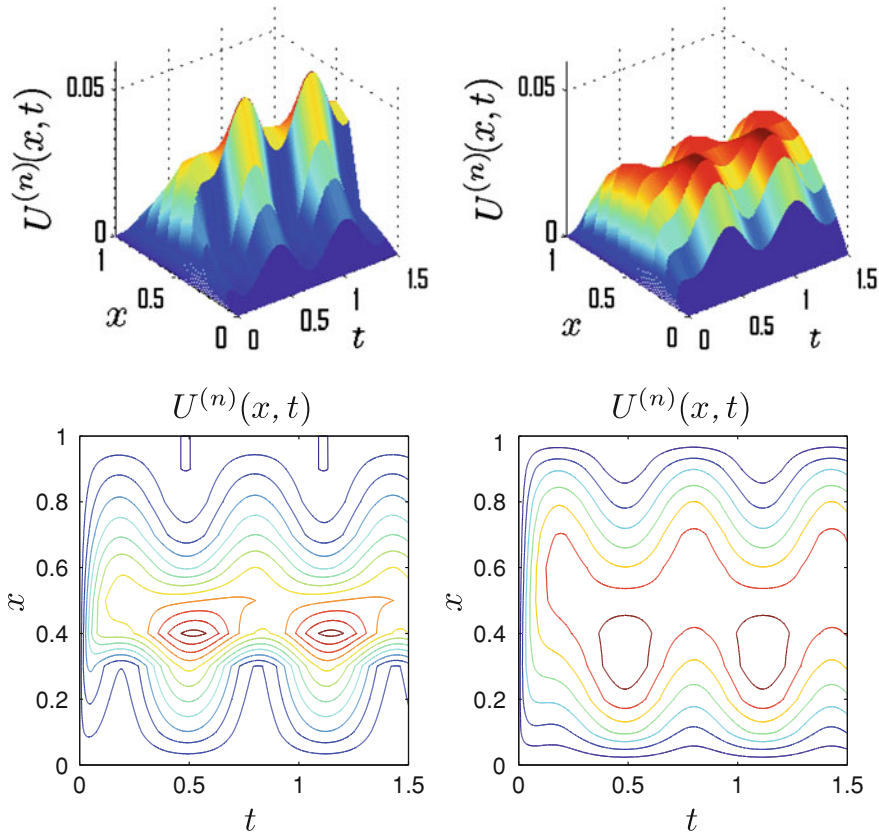
Let  $\bar{v} > 0$  be a sufficiently large frequency such that  $\int_0^{\bar{v}} g(v) dv \simeq 1$  and set  $G^{(n)}(x) = \sum_{k=1}^n \sigma_k^{(n)} (A_k \cos(v_k x) + B_k \sin(v_k x))$ , where  $v_k = (k - 1/2) \bar{v}/n$ ,  $A_k$ ,



**Fig. 9.3** A sample of  $U^{(n)}(x, t)$  corresponding to a sample  $\Sigma^{(n)}(x, \omega)$  of  $\Sigma^{(n)}(x)$  (top left panel) and the spatial average of  $\Sigma^{(n)}(x, \omega)$  (top right panel). The bottom panels are contour lines of the plots in the top panels

$B_k$  are independent  $N(0, 1)$  variables, and  $\sigma_k^{(n)}$  is the square root of the area under  $g(v)$  in frequency band  $(v_k - \bar{v}/(2n), v_k + \bar{v}/(2n))$ . Under mild conditions as  $n \rightarrow \infty$  and  $\bar{v} \rightarrow \infty$ ,  $G^{(n)}(x)$  and  $\Sigma^{(n)}(x)$  become versions of  $G(x)$  and  $\Sigma(x)$ , respectively,  $\Sigma^{(n)}(x)$  converges to  $\Sigma(x)$  in the mean square sense for each  $x \in (0, l)$ , and  $U^{(n)}(x, t)$  converges in mean square to  $U(x, t)$  in any bounded time interval.

The plots in Figs. 9.3 and 9.4 are for  $l = 1$ , a Beta distribution  $F$  with range  $(a = 0.01, b = 6.0)$  and shape parameters  $\gamma = \eta = 1$ ,  $\lambda = 10$ , and a finite difference scheme with space and time steps  $\Delta x = l/10$  and  $\Delta t = 1.5/2000$ . The parametric model  $G^{(n)}(x)$  for  $G(x)$  has been obtained by truncating the spectral density of  $G(x)$  at  $\bar{v} = 30$ , scaling the truncated spectral density to have unit area, and approximating it by a discrete spectral density with  $n = 100$  equally spaced frequencies. The parametric model for conductivity is  $\Sigma^{(n)}(x) = F^{-1} \circ \Phi(G^{(n)}(x))$ . The top left and right panels in Fig. 9.3 show a sample of  $U^{(n)}(x, t)$  corresponding to a sample  $\Sigma^{(n)}(x, \omega)$  of



**Fig. 9.4** A sample of  $U^{(n)}(x, t)$  corresponding to a sample  $\Sigma^{(n)}(x, \omega)$  of  $\Sigma^{(n)}(x)$  (top left panel) and the spatial average of  $\Sigma^{(n)}(x, \omega)$  (top right panel). The bottom panels are contour lines of the plots in the top panels

$\Sigma^{(n)}(x)$  and a sample of  $U^{(n)}(x, t)$  corresponding to the spatial average  $\bar{\Sigma}^{(n)}(x, \omega) = (1/l) \int_0^l \Sigma^{(n)}(x, \omega) dx$  of conductivity sample  $\Sigma^{(n)}(x, \omega)$ . The bottom panels in the figure are contour lines of the three-dimensional plots in the top panels. Figure 9.4 provides the same information for another samples of  $\Sigma^{(n)}(x)$ . The plots show that  $U^{(n)}(x, t)$  exhibits a notable sample-to-sample variation and the solution of (9.60) cannot be approximated by that of this equation corresponding to a homogeneous specimen with properties given by the spatial average of  $\Sigma^{(n)}(x)$ .  $\diamond$

*Proof* That  $G^{(n)}(x)$  becomes a version of  $G(x)$  as  $n \rightarrow \infty$  and  $\bar{v} \rightarrow \infty$  follows from Theorem 3.46. Since

$$\begin{aligned} P(\cap_{i=1}^m \{\Sigma^{(n)}(x_i) \leq z_i\}) &= P(\cap_{i=1}^m \{G^{(n)}(x_i) \leq \Phi^{-1} \circ F(z_i)\}) \\ &\rightarrow P(\cap_{i=1}^m \{G(x_i) \leq \Phi^{-1} \circ F(z_i)\}) = P(\cap_{i=1}^m \{\Sigma(x_i) \leq z_i\}), \quad n \rightarrow \infty, \quad \bar{v} \rightarrow \infty, \end{aligned}$$

holds for any integer  $m \geq 1$ , arbitrary arguments  $x_i$ , and  $z_i \in \mathbb{R}$  by properties of  $G^{(n)}(x)$ ,  $\Sigma^{(n)}(x)$  becomes a version of  $\Sigma(x)$  as  $n \rightarrow \infty$  and  $\bar{v} \rightarrow \infty$ .

If  $F^{-1} \circ \Phi$  is Lipschitz continuous, we have

$$|\Sigma^{(n)}(x) - \Sigma(x)| = |F^{-1} \circ \Phi(G^{(n)}(x)) - F^{-1} \circ \Phi(G(x))| \leq c |G^{(n)}(x) - G(x)|, \quad (9.61)$$

where  $c > 0$  is a constant. This implies  $E[|\Sigma^{(n)}(x) - \Sigma(x)|^2] \leq c^2 E[|G^{(n)}(x) - G(x)|^2]$  so that  $E[|\Sigma^{(n)}(x) - \Sigma(x)|^2] \rightarrow 0$  by the mean square convergence of  $G^{(n)}(x)$  to  $G(x)$ , so that  $\Sigma^{(n)}(x)$  converges in the mean square sense to  $\Sigma(x)$ .

Let  $(\xi_1, \dots, \xi_d)$  be discrete points in  $(0, l)$  defining the nodes of a finite difference mesh used to approximate the spatial derivatives of  $U(x, t)$ . The  $\mathbb{R}^d$ -valued stochastic process  $Y(t) = (Y_1(t) = U(\xi_1, t), \dots, Y_d(t) = U(\xi_d, t))'$  satisfies the ordinary differential equation  $\dot{Y}(t) = A Y(t) + W(t)$ ,  $t \geq 0$ , where  $A$  denotes a  $(d, d)$  random matrix that incorporates the boundary conditions for (9.60) and  $W(t) = (W_1(t) = V(\xi_1, t), \dots, W_d(t) = V(\xi_d, t))'$ . Consider also the  $\mathbb{R}^d$ -valued process  $Y^{(n)}(t)$  defined by  $\dot{Y}^{(n)}(t) = A^{(n)} Y^{(n)}(t) + W(t)$  in which  $A^{(n)}$  is  $A$  with  $\Sigma^{(n)}$  in place of  $\Sigma$ . If the largest eigenvalue  $\Lambda_{\max}$  of  $(A + A')/2$  is such that  $P(\Lambda_{\max} < 0) = 1$ , the discrepancy between  $Y^{(n)}(t)$  and  $Y(t)$  depends on that between  $A$  and  $A^{(n)}$ , and can be bounded by

$$\|Y(t) - Y^{(n)}(t)\| \leq e^{\Lambda_{\max} t} \int_0^t e^{-\Lambda_{\max} s} \|(A - A^{(n)}) Y^{(n)}(s)\| ds,$$

where  $\|\cdot\|$  is the Euclidean norm (Theorem 7.28). If there exists  $M > 0$  such that  $\|Y^{(n)}(t)\| \leq M$  at all times a.s., we have

$$\begin{aligned} \|Y(t) - Y^{(n)}(t)\| &\leq \|A - A^{(n)}\| e^{\Lambda_{\max} t} \int_0^t e^{-\Lambda_{\max} s} \|Y^{(n)}(s)\| ds \\ &\leq \frac{M \|A - A^{(n)}\|}{-\Lambda_{\max}} (1 - e^{\Lambda_{\max} t}) \leq \frac{M \|A - A^{(n)}\|}{-\Lambda_{\max}} \end{aligned}$$

since  $\|(A - A^{(n)}) Y^{(n)}(s)\| \leq \|A - A^{(n)}\| \|Y^{(n)}(s)\|$ . This gives  $E[\|Y(t) - Y^{(n)}(t)\|^2] \leq (M/\alpha)^2 E[\|A - A^{(n)}\|^2]$ , where  $\alpha > 0$  is such that  $P(|\Lambda_{\max}| > \alpha) = 1$ . Since  $E[\|A - A^{(n)}\|^2]$  is a sum of expectations  $E[(\Sigma(\xi_p) - \Sigma^{(n)}(\xi_p))(\Sigma(\xi_q) - \Sigma^{(n)}(\xi_q))]$  with absolute value smaller than

$$\left( E[(\Sigma(\xi_p) - \Sigma^{(n)}(\xi_p))^2] E[(\Sigma(\xi_q) - \Sigma^{(n)}(\xi_q))^2] \right)^{1/2}$$

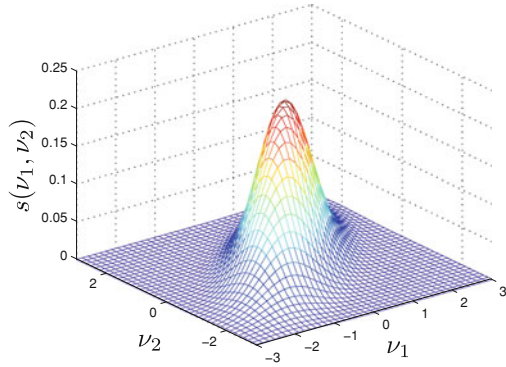
by the Cauchy–Schwarz inequality, we have  $E[\|A - A^{(n)}\|^2] \rightarrow 0$  as  $n \rightarrow \infty$ , which implies the mean square convergence of  $Y^{(n)}(t)$  to  $Y(t)$ .  $\blacktriangle$

We have seen that Monte Carlo solutions of SPDEs require the representation of all random elements in these equations by parametric models. Parametric models used by the stochastic Galerkin and collocation methods are discussed later in this chapter. The models used by these methods are required to satisfy additional conditions that



**Fig. 9.5** Spectral density

$s(v)$  of  $G(x)$  for  
 $v \in [-3, 3] \times [-3, 3]$



may limit their value in applications. The following example explores effects of these conditions on the second moment properties of the solution of a stochastic elliptic boundary value problem.

*Example 9.7* Suppose  $a(x)$ ,  $x \in D = (0, l_1) \times (0, l_2) \subset \mathbb{R}^2$ , in (9.53) is a Beta translation random field defined by

$$a(x) = \alpha + (\beta - \alpha) F_{\text{Beta}(p,q)}^{-1} \circ \Phi(G(x)) = h(G(x)), \quad x \in D, \quad (9.62)$$

where  $F_{\text{Beta}(p,q)}$  denotes the distribution of a standard Beta random variable with shape parameters  $(p, q)$ , mean  $p/(p+q)$ , and variance  $pq/[(p+q)^2(p+q+1)]$ , and  $0 < \alpha < \beta < \infty$  are constants. Note that  $a(x)$  takes values in the interval  $[\alpha, \beta]$ , so that (9.53) is an elliptic partial differential equation a.s. The image  $G(x)$  of  $a(x)$  is a homogeneous Gaussian field with mean 0, variance 1, spectral density

$$s(v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{v_1^2 - 2\rho v_1 v_2 + v_2^2}{2(1-\rho^2)}\right], \quad v = (v_1, v_2) \in \mathbb{R}^2, \quad (9.63)$$

and covariance function

$$c(\tau) = E[G(x)G(x+\tau)] = \exp\left[-\frac{\tau_1^2 + 2\rho\tau_1\tau_2 + \tau_2^2}{2}\right], \quad \tau = (\tau_1, \tau_2) \in \mathbb{R}^2, \quad (9.64)$$

where  $\rho \in (-1, 1)$ . Figure 9.5 shows the spectral density of  $G(x)$  for  $\rho = 0.7$  and  $v = (v_1, v_2)$  in the rectangle  $[-3, 3] \times [-3, 3]$ . Samples of  $a(x)$  can be calculated from samples of  $G(x)$  that can be generated by various Monte Carlo algorithms (Sect. 3.8). The second moment properties of  $a(x)$  are  $E[a(x)] = \alpha + (\beta - \alpha)p/(p+q)$  and  $c_a(\tau) = E[(a(x+\tau) - E[a(x)])(a(x) - E[a(x)])] = E[(h(G(x+\tau)) - E[a(x)])(h(G(x)) - E[a(x)])]$ . Translation random fields of the type in (9.62) can be used to describe the spatial variability of conductivity or other properties of random microstructures.

Current implementations of the stochastic Galerkin and collocation methods approximate  $a(x)$  by linear parametric models  $a^{(n)}(x)$ , that is, finite sums of specified



deterministic functions of  $x \in D$  with random coefficients (Sects. 9.4.8 and 9.4.9). Truncated Karhunen–Loève series or discrete spectral representations are commonly used to construct linear parametric models. For example, the linear parametric model

$$a^{(n)}(x) = a_0 + \sum_{k=1}^n \sigma_k (C_k \cos(v^{(k)} \cdot x) + D_k \sin(v^{(k)} \cdot x)) \quad (9.65)$$

constitutes a discrete spectral representation of  $a(x)$ , where  $\{\sigma_k\}$  and  $a_0$  are constants,  $v^{(k)} \in \mathbb{R}^2$  denote frequencies, and  $\{C_k, D_k\}$  are uncorrelated random variables with mean 0 and variance 1. The mean and covariance functions of the parametric model  $a^{(n)}(x)$  are  $E[a^{(n)}(x)] = a_0$  and  $c_a^{(n)}(\tau) = E[(a^{(n)}(x + \tau) - a_0)(a^{(n)}(x) - a_0)] = \sum_{k=1}^n \sigma_k^2 \cos(v^{(k)} \cdot \tau)$ .

The model size  $n$  in (9.65) defines the dimension of the probability space, that is, the number of random variables in the expression of  $a^{(n)}(x)$ , and is selected based on computational considerations. The frequencies  $\{v^{(k)}\}$  can be chosen at the centers of rectangles  $\{I_k\}$  partitioning the frequency band  $I = [-\bar{v}, \bar{v}]^2$ ,  $0 < \bar{v} < \infty$ , of the spectral density  $s_a(v)$  of  $a(x)$ , that is,  $I = \cup_{k=1}^n I_k$  and  $I_k \cap I_l = \emptyset$  for  $k \neq l$ , where  $I$  is such that  $\int_I s_a(v) dv \simeq 1$ . The second moment properties of  $a^{(n)}(x)$  with  $\sigma_k^2 = \int_{I_k} s_a(v) dv$  converge to those of  $a(x)$  as  $n \rightarrow \infty$  and  $\bar{v} \rightarrow \infty$  (Sect. 3.8.1.2), so that  $a^{(n)}(x)$  with  $a_0 = E[a(x)]$  is approximately equal to  $a(x)$  in the second moment sense for sufficiently large  $n$  and  $\bar{v}$ .

However, the samples of  $a^{(n)}(x)$  and  $a(x)$  may have very different properties. For example,  $a(x)$  takes values in  $[\alpha, \beta]$  while the samples of  $a^{(n)}(x)$  may leave this interval with a non-zero probability that depends on the distributions postulated for  $\{C_k, D_k\}$ . To confine the samples of  $a^{(n)}(x)$  to a bounded interval in  $(0, \infty)$  so that (9.53) with  $a^{(n)}(x)$  in place of  $a(x)$  is elliptic, it is commonly assumed in the stochastic Galerkin and collocation methods that the distributions of  $\{C_k, D_k\}$  have bounded support, that depends on  $n$  and are such that  $a^{(n)}(x)$  has strictly positive samples. A version of this model is the random field

$$a^{(n,\zeta)}(x) = a_0 + \zeta \sum_{k=1}^n \sigma_k (C_k \cos(v^{(k)} \cdot x) + D_k \sin(v^{(k)} \cdot x)), \quad (9.66)$$

which resembles  $a^{(n)}(x)$  except that its random part is scaled by a factor  $0 < \zeta < 1$  such that  $\min_{x \in D} \{a^{(n,\zeta)}(x)\} > 0$  a.s. The factor  $\zeta$  depends on  $n$  and the distributions of  $\{C_k, D_k\}$ . Note that  $a^{(n,\zeta)}(x)$  with  $a_0 = E[a(x)]$  has the same mean as  $a(x)$  and that the covariance function of this field scaled by  $\zeta^2$ , that is, the function  $c_a^{(n)}(\tau)/\zeta^2$ , converges to  $c_a(\tau)$  as  $n \rightarrow \infty$  and  $\bar{v} \rightarrow \infty$ , so that the approximation  $c_a^{(n)}(\tau)/\zeta^2 \simeq c_a(\tau)$  is satisfactory for sufficiently large values of  $n$  and  $\bar{v}$ . The variance of  $a^{(n,\zeta)}(x)$  can be much smaller than that of the target random field  $a(x)$  since  $\text{Var}[a^{(n,\zeta)}(x)] \simeq \zeta^2 \text{Var}[a(x)]$ ; it approaches 0 as  $n$  increases indefinitely since  $a^{(n)}(x)$  with independent random variables  $\{C_k, D_k\}$  of bounded support becomes

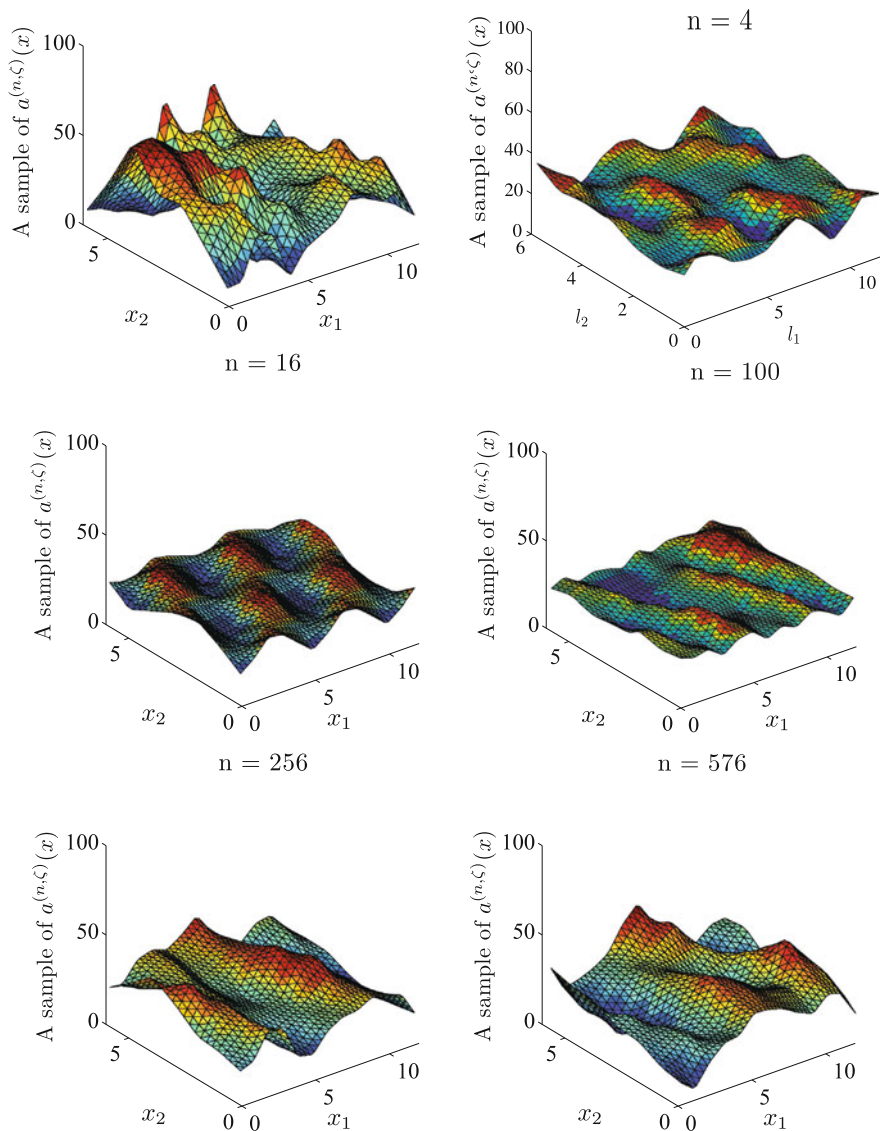
a Gaussian field as  $n \rightarrow \infty$  (Sect. 6.3.3.1). Moreover, the distribution of  $a^{(n,\zeta)}(x)$  depends on  $n$  and differs from that of  $a(x)$ .

In summary, the model  $a^{(n,\zeta)}(x)$  of  $a(x)$  in (9.66) can be tuned to have the same mean as  $a(x)$ . Its scaled covariance function  $c_a^{(n)}(\tau)/\zeta^2$  approximates satisfactorily the covariance function  $c_a(\tau)$  of  $a(x)$  for adequate values of  $n$  and  $\bar{v}$ . The variance of  $a^{(n,\zeta)}(x)$  can be much smaller than that of  $a(x)$  since  $\text{Var}[a^{(n,\zeta)}(x)] \simeq \zeta^2 \text{Var}[a(x)]$ . For example, if  $\{C_k, D_k\}$  take values in  $(-\beta_k, \beta_k)$ ,  $0 < \beta_k < \infty$ ,  $k = 1, \dots, n$ , then  $|a^{(n,\zeta)}(x) - a_0| \leq \zeta \sum_{k=1}^n \sigma_k (C_k^2 + D_k^2)^{1/2} \leq \sqrt{2} \zeta \sum_{k=1}^n \sigma_k \beta_k$  so that  $a^{(n,\zeta)}(x)$  has strictly positive samples if  $a_0 - \sqrt{2} \zeta \sum_{k=1}^n \sigma_k \beta_k > 0$  or  $\zeta < a_0 / (\sqrt{2} \sum_{k=1}^n \sigma_k \beta_k)$ . Also, the distributions of  $a(x)$  and  $a^{(n,\zeta)}(x)$  differ.  $\diamond$

*Example 9.8* Let  $U(x)$  denote the temperature field in a rectangular specimen satisfying the stochastic partial differential equation (9.53) with  $f(x) = 0$ ,  $D = (0, l_1) \times (0, l_2)$ ,  $l_1 = 12$ ,  $l_2 = 6$ ,  $U(0, x_2) = 0$  and  $U(l_1, x_2) = 1$  for  $x_2 \in (0, l_2)$ , and  $\partial U(x_1, 0)/\partial x_2 = \partial U(x_1, l_2)/\partial x_2 = 0$  for  $x_1 \in (0, l_1)$ . The conductivity field  $a(x)$  is given by (9.62), (9.63), and (9.64) with  $\alpha = 0.25$ ,  $\beta = 100$ ,  $p = 1$ ,  $q = 3$ ,  $\rho = 0.7$ , and  $\bar{v} = 3$ . The coordinates of the frequencies  $v^{(k)}$  in (9.65) are  $v_r^{(k)} = -\bar{v} + (k - 1/2)(2\bar{v}/n)$ ,  $r = 1, 2$ . Since the difference between the covariance function  $c(\tau)$  of  $G(x)$  and the scaled covariance  $c_a(\tau)/c_a(0)$  of  $a(x)$  is small ([22], Sect. 3.1.1), we assume  $c_a(\tau)/c_a(0) \simeq c(\tau)$ .

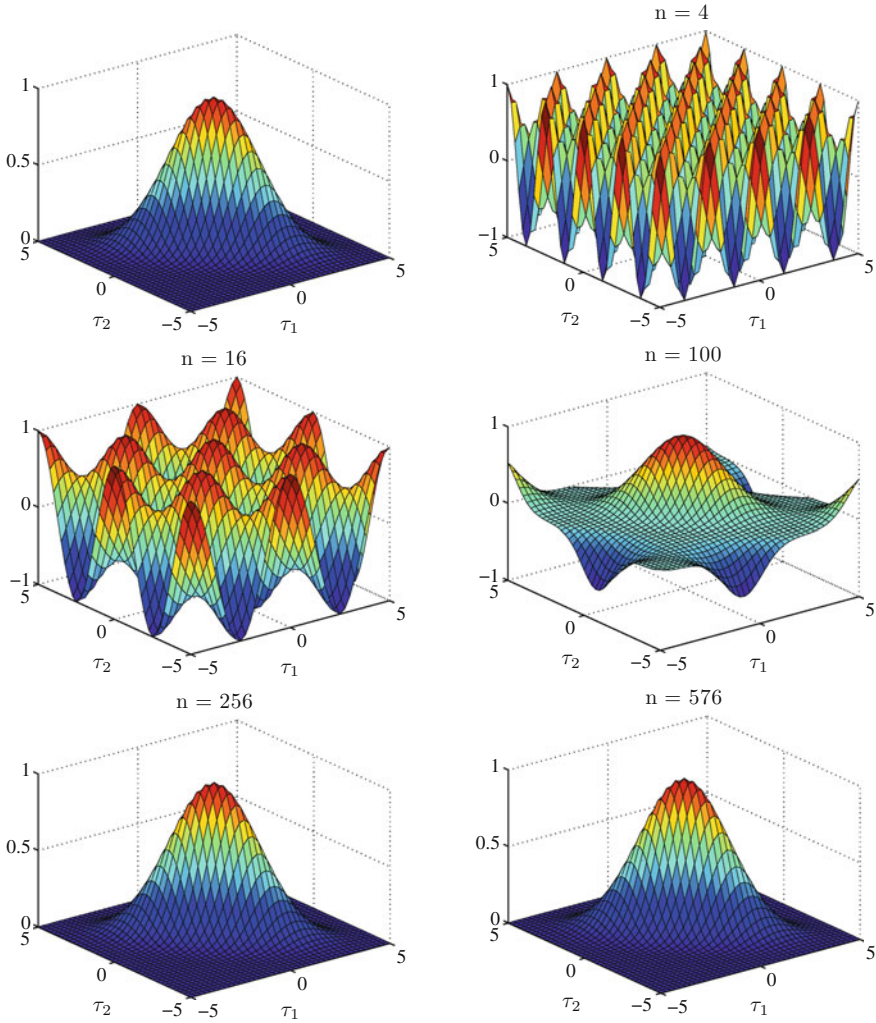
Let  $U^{(n,\zeta)}(x)$  be the solution of (9.53) with  $a^{(n,\zeta)}(x)$  in place of  $a(x)$ . The random variables in the definition of  $a^{(n,\zeta)}(x)$  are assumed to be independent, uniformly distributed with mean 0 and variance 1, that is,  $C_k, D_k \sim U(-\sqrt{3}, \sqrt{3})$ ,  $k = 1, \dots, n$ . The scale factor  $\zeta$  is selected such that less than approximately 5% of the samples of  $a^{(n,\zeta)}(x)$  take negative value in  $D$ , for example,  $\zeta = 0.3286$  for  $n = 4$ . The samples of  $a^{(n,\zeta)}(x)$  taking negative values in  $D$  are not used to calculate samples of  $U^{(n,\zeta)}(x)$ , so that the reported estimates are for the random field  $U^{(n,\zeta)}(x) \mathbf{1}(\min_{x \in D} \{a^{(n,\zeta)}(x)\} > 0)$ .

Properties of solutions  $U(x)$  and  $U^{(n,\zeta)}(x)$  are estimated from samples of these fields obtained by solving (9.53) with  $f(x) = 0$  for 800 independent samples of  $a(x)$  and  $a^{(n,\zeta)}(x)$ , respectively. Figure 9.6 shows a sample of the Beta translation conductivity  $a(x)$  (top left panel) and samples of  $a^{(n,\zeta)}(x)$  for several values of  $n$ . The skewness and kurtosis coefficients of  $a(x)$  are 0.8589 and 3.0952 ([23], Chap. 24). Their estimates obtained from 800 samples are 0.9546 and 3.4143. Estimates of skewness and kurtosis coefficients for  $a^{(n,\zeta)}(x)$  based also on 800 independent samples of this field are 0.0602 and 2.4714 for  $n = 4$ , and  $-0.0797$  and 2.7855 for  $n = 576$ . While the density of  $a(x)$  is skewed, the density of  $a^{(n,\zeta)}(x)$  is approximately symmetric about its mean. Figure 9.7 shows the covariance functions of  $a(x)$  (top left panel) and  $a^{(n,\zeta)}(x)$  for several values of  $n$ . The covariance function of  $a^{(n,\zeta)}(x)$  coincides with that of  $(a^{(n,\zeta)}(x) - a_0)/\zeta$  and approaches the covariance function of  $a(x)$  as  $n$  increases, in agreement with observations in Example 9.7. Figure 9.8 shows the expectation of  $U(x)$  (top left panel) and samples of  $U^{(n,\zeta)}(x)$  for  $n = 4, 16, 100, 256$ , and 576. The samples of  $U^{(n,\zeta)}(x)$  for these values of  $n$  nearly coincide with  $E[U(x)]$ . Figure 9.9 shows the standard deviation of  $U(x)$  and the standard deviations of  $U^{(n,\zeta)}(x)$  for



**Fig. 9.6** A sample of  $a(x)$  (top left panel) and samples of  $a^{(n, \zeta)}(x)$  for  $n = 4, 16, 100, 256$ , and  $576$

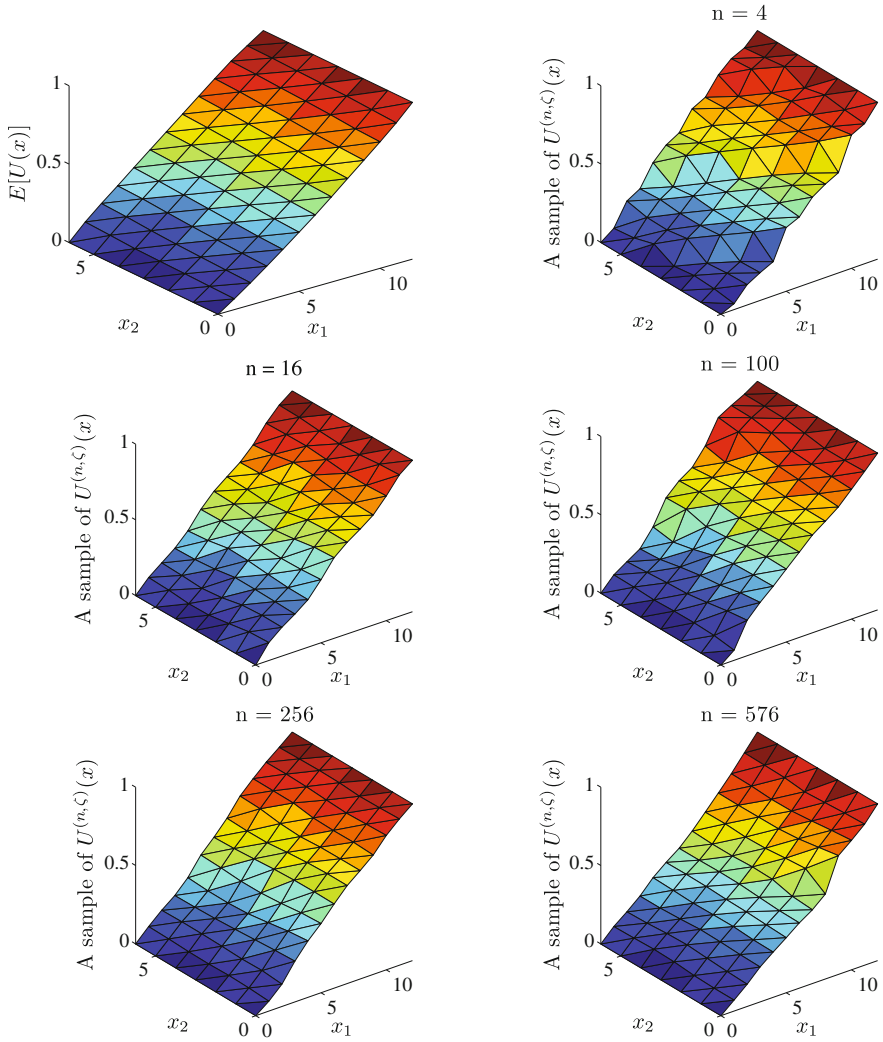
$n = 4, 16, 100, 256$ , and  $576$ . The standard deviation of  $U^{(n, \zeta)}(x)$  is much smaller than that of  $U(x)$  for all values of  $n$  in the figure. The dependence of the spatial variation of the standard deviation of  $U^{(n, \zeta)}(x)$  on  $n$  is consistent with the facts that the probability law of  $a^{(n, \zeta)}(x)$  is a function of  $n$  and the mapping  $a^{(n, \zeta)} \mapsto U^{(n, \zeta)}$  is nonlinear.



**Fig. 9.7** Covariance functions of  $a(x)$  (top left panel) and  $a^{(n)}(x)$  for  $n = 4, 16, 100, 256$ , and  $576$

The discrepancy between the solutions  $U(x)$  and  $U^{(n,\xi)}(x)$  can be bounded by (9.57) with  $\|a - \tilde{a}\|_{L^\infty(D \times \Omega)} = \|a - a^{(n,\xi)}\|_{L^\infty(D \times \Omega)}$  and  $\|f - \tilde{f}\|_{L^2(D \times \Omega)} = 0$ . The bound remains strictly positive for all value of  $n$  since  $\|a - a^{(n,\xi)}\|_{L^\infty(D \times \Omega)}$  does not converge to 0 as  $n \rightarrow \infty$ .  $\diamond$

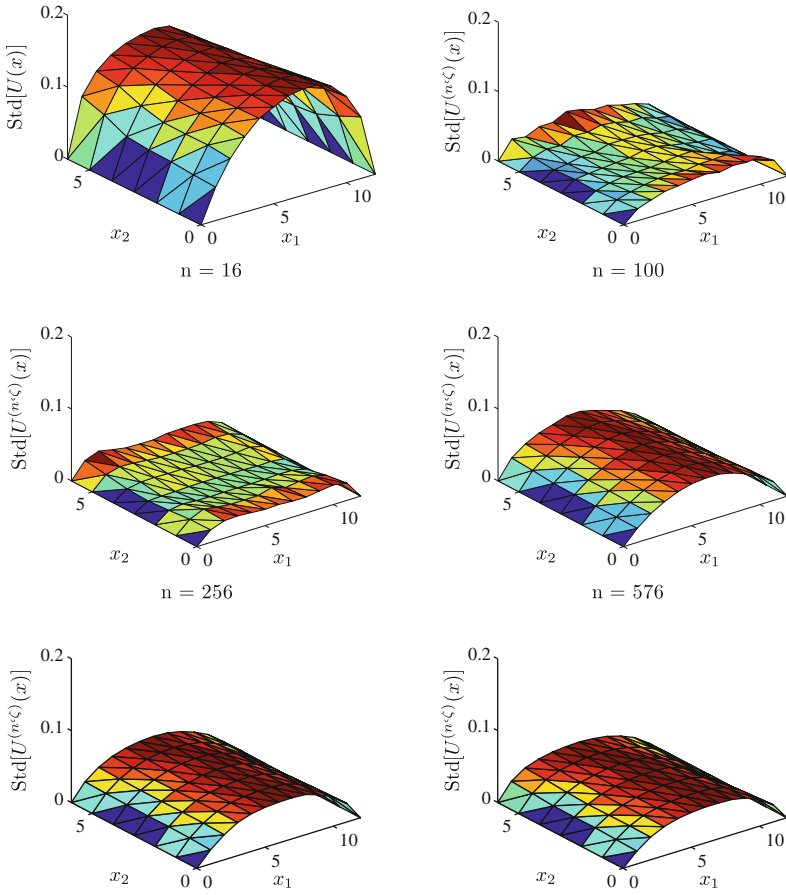
The stochastic partial differential equation in this example is also solved in Examples 9.12, 9.13, 9.16, and 9.19 by the stochastic reduced order models, stochastic Galerkin, and stochastic collocation methods to illustrate the implementation of these methods and assess their relative accuracy and computational efficiency.



**Fig. 9.8** Expectation of  $U(x)$  (top left panel) and samples of  $U^{(n,\zeta)}(x)$  for  $n = 4, 16, 100, 256,$  and  $576$

We conclude this section with the observation that PDEs can be viewed as filters whose solutions to random input are random fields. The spectral properties of these fields depend on the input probability law and the functional form of the PDE under consideration. These stochastic equations can be used to generate samples of Gaussian random fields with specified spectral properties.

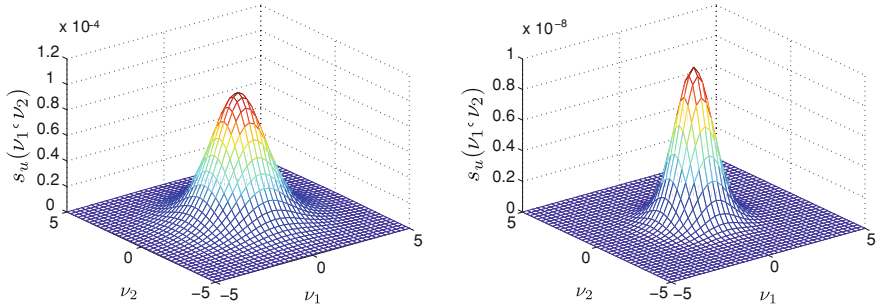
**Example 9.9** Consider the SPDE  $\mathcal{L}[U(x)] = V(x)$ ,  $x \in \mathbb{R}^d$ , in (9.40) and suppose  $\mathcal{L} = (\Delta - \alpha^2)^{(2p)}$  and  $V(x)$  is a weakly homogeneous random field with mean 0 and spectral density  $s(\nu) = 1(\nu \in R)$ , where  $\alpha > 0$  is a constant,  $p \geq 1$



**Fig. 9.9** Standard deviations of  $U(x)$  (top left panel) and  $U^{(n,\xi)}(x)$  for  $n = 4, 16, 100, 256$ , and  $576$

is an integer,  $\Delta$  denotes the Laplace operator, the superscript  $(2p)$  indicates that the operator in parenthesis is applied  $2p$  times, and  $R = [-\bar{v}, \bar{v}]^d$ ,  $0 < \bar{v} < \infty$ . The solution of this equation to  $V(x) = e^{i v \cdot x}$  is  $e^{i v \cdot x} / (\|v\|^2 + \alpha^2)^{2p}$ , so that  $U(x)$  has mean zero and spectral density  $s_u(v) = 1(v \in R) / (\|v\|^2 + \alpha^2)^{2p}$  ([24], Theorems 2.1.2 and 2.4.1, and Sect. 3.5.3 in this book). Figure 9.10 shows spectral densities of  $U(x)$  for  $d = 2$ ,  $\bar{v} = 5$ ,  $\alpha = 10$ ,  $p = 1$  (left panel), and  $p = 2$  (right panel). Alternative spectral densities can be obtained for  $U(x)$  by using other functional forms for  $\mathcal{L}$ .  $\diamond$





**Fig. 9.10** Spectral densities of  $U(x)$  for  $\alpha = 10$ ,  $p = 1$  (left panel), and  $p = 2$  (right panel)

### 9.4.5 Stochastic Reduced Order Model Method

We construct approximations  $\tilde{U}(x, t)$  for the solution of  $U(x, t)$  of (9.33) by representing the random entries in the definition of this equation by simple random elements, referred to as stochastic reduced order models (SROMs). Two SROM-based methods are presented for solving stochastic partial differential equations. The first method approximates the solution  $U$  by a SROM  $\tilde{U}$  derived from a SROM of the random entries in (9.33). The second method represents the dependence of  $U$  on the random entries of (9.33) by piecewise linear functions. The construction of this representation is less simple than that in the first method, but provides a more accurate approximation for  $U$ . We refer to the first and second methods as SROM- and extended stochastic reduced order model (ESROM)-based solutions. The implementation of the first method is discussed in Sect. 9.4.6. The method is applied to solve initial and boundary value problems (Sects. 9.4.6.1 and 9.4.6.2). The implementation of the second method is discussed in Sect. 9.4.7. An example is used to illustrate the application of the method and assess its accuracy.

### 9.4.6 Stochastic Reduced Order Models

We construct a SROM  $\tilde{U}(x, t)$  for the solution  $U(x, t)$  of (9.33) and use it to find properties of  $U(x, t)$  approximately. The construction involves three steps. First, a SROM  $\tilde{X}$  is developed for the random entries in (9.33) that are collected in a random element  $X$ . Second, a SROM  $\tilde{U}(x, t)$  is constructed for  $U(x, t)$  from the solutions of (9.33) with  $X$  replaced by the samples of  $\tilde{X}$  and the probabilities of these samples. Third, properties of  $U(x, t)$  are approximated by those of  $\tilde{U}(x, t)$ .

Let  $X$  be a random element defined on a probability space  $(\Omega, \mathcal{F}, P)$  with values in a metric space  $(S, d)$ , where  $d$  denotes a metric on  $S$ . For example,  $X$  is a real-valued random variable, a random vector, or a real-valued random function with continuous

samples defined on  $[0, 1]$  if  $S$  is the real line  $\mathbb{R}$ , the Euclidean space  $\mathbb{R}^{d'}$ , or the space of real-valued continuous function  $C[0, 1]$  defined on  $[0, 1]$ .

A SROM  $\tilde{X}$  for  $X$  is a simple random element with  $m$  distinct outcomes  $\tilde{x} = (\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)})$  occurring with probabilities  $p = (p_1, \dots, p_m)$ , that is,  $P(\tilde{X} = \tilde{x}^{(k)}) = p_k$ , where  $p_k \geq 0$ ,  $k = 1, \dots, m$ , and  $\sum_{k=1}^m p_k = 1$ . Any set  $\tilde{x}$  of distinct points in  $S$  can be used for the range of  $\tilde{X}$ . However, it is unlikely that the probability law of  $\tilde{X}$  corresponding to an arbitrary selection of  $\tilde{x}$  and  $p$  will be similar to that of  $X$ . We are interested in SROMs  $\tilde{X}$  with size  $m$  that are optimal in some sense (Sect. A.3). We denote the samples of any random element  $W$  by  $\{w^{(i)}\}$  or  $\{w_i\}$ . Similarly, the samples of a SROM  $\tilde{W}$  for  $W$  are denoted by  $\{\tilde{w}^{(k)}\}$  or  $\{\tilde{w}_k\}$ .

Suppose the size  $m$  and the range  $(\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)})$  of a SROM  $\tilde{X}$  have been selected and that our objective is to find optimal values for  $p = (p_1, \dots, p_m)$ . The model size  $m$  is essentially determined by the computational effort required to solve  $m$  deterministic versions of (9.33) obtained by setting  $X$  equal to samples of  $\tilde{X}$ . The discrepancy between  $\tilde{X}$  and  $X$  can be measured by, for example, the error  $\sum_{q \geq 1} \alpha_q e_q(p)$ , where  $\alpha_q > 0$  are weighting factors and  $e_q(p)$  measure differences between various properties of  $X$  and  $\tilde{X}$ . For example, if  $X$  is a real-valued random variable with finite moments  $\mu_r = E[X^r]$ ,  $r = 1, 2, \dots, \bar{r}$ , and distribution  $F$ , then we may set  $e_1(p, \tilde{x}) = \max_x |F(x) - \tilde{F}(x)|$  or  $e_1(p, \tilde{x}) = \int (F(x) - \tilde{F}(x))^2 w(x) dx$  and  $e_2(p, \tilde{x}) = \sum_{r=1}^{\bar{r}} w_r (\mu_r - \tilde{\mu}_r)^2$ , where  $w(x) \geq 0$  is an integrable function,  $w_r > 0$  are weighting factors,  $\tilde{\mu}_r = \sum_{k=1}^m (\tilde{x}^{(k)})^r p_k$ , and  $\tilde{F}(x) = \sum_{k=1}^m 1(\tilde{x}^{(k)} \leq x) p_k$ . The defining parameters  $\tilde{x} = (\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)})$  and  $p = (p_1, \dots, p_m)$  for a SROM  $\tilde{X}$  with size  $m$  are such that they minimize the objective function

$$e(p, \tilde{x}) = \sum_{q \geq 1} \alpha_q e_q(p, \tilde{x}) \quad (9.67)$$

under the constraints  $p_k \geq 0$ ,  $k = 1, \dots, m$ , and  $\sum_{k=1}^m p_k = 1$ .

We present an algorithm for finding optimal values of  $p$  for specified samples  $\tilde{x} = (\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)})$  of  $\tilde{X}$ , that is, the vector  $p$  minimizing the objective function

$$e(p) = \sum_{q \geq 1} \alpha_q e_q(p) \quad (9.68)$$

under the same constraint as in (9.67), where the errors  $e_q(p)$  are equal to  $e_q(p, \tilde{x})$  for a specified sample  $\tilde{x}$ . Suppose  $n_{\text{set}}$  sets of  $m$  independent samples of  $X$  have been generated and let  $p^{\text{opt}, j}$  denote the optimal probability vector for set  $j = 1, \dots, n_{\text{set}}$ , that is, the vector minimizing the objective function for this set. The SROM  $\tilde{X}$  corresponds to the set of  $m$  samples that has the minimum optimal error, that is,  $\tilde{X}$  has the range of set  $j_0 \in \{1, 2, \dots, n_{\text{set}}\}$  and probability vector  $p^{\text{opt}, j_0}$  if  $e(p^{\text{opt}, j_0}) \leq e(p^{\text{opt}, j})$ ,  $j = 1, \dots, n_{\text{set}}$ . Alternative algorithms for constructing SROMs can be found in [19].

Suppose a SROM  $\tilde{X}$  with defining parameters  $(\tilde{x}^{(k)}, p_k)$ ,  $k = 1, \dots, m$ , has been selected for the random elements in (9.33). Let  $\tilde{u}^{(k)}(x, t)$ ,  $k = 1, \dots, m$ ,



be solutions of (9.33) for  $X$  set equal to the samples  $\tilde{x}^{(k)}$ ,  $k = 1, \dots, m$ , of  $\tilde{X}$ . The solutions  $(\tilde{u}^{(1)}(x, t), \dots, \tilde{u}^{(m)}(x, t))$  and the probabilities  $(p_1, \dots, p_m)$  of the samples of  $\tilde{X}$  define a SROM  $\tilde{U}(x, t)$  for  $U(x, t)$ . Any deterministic solver can be used to find  $\tilde{u}^{(k)}(x, t)$ ,  $k = 1, \dots, m$ . Properties of  $\tilde{U}(x, t)$  can be calculated simply. For example, the marginal moments of order  $r \geq 1$ , the marginal distribution, and the correlation function of  $\tilde{U}(x, t)$  are

$$\begin{aligned} E[\tilde{U}(x, t)^r] &= \sum_{k=1}^m (\tilde{u}^{(k)}(x, t))^r p_k, \\ P(\tilde{U}(x, t) \leq \xi) &= \sum_{k=1}^m 1(\tilde{u}^{(k)}(x, t) \leq \xi) p_k, \text{ and} \\ E[\tilde{U}(x, s) \tilde{U}(y, t)] &= \sum_{k=1}^m \tilde{u}^{(k)}(x, s) \tilde{u}^{(l)}(y, t) p_k, \end{aligned} \quad (9.69)$$

and can be used to approximate properties of  $U(x, t)$ .

The objective function of the optimization algorithm in (9.68) quantifies differences between global properties of  $X$  and  $\tilde{X}$  so that these random elements may or may not be defined on the same probability space. Similarly, comparisons between global properties of the solutions  $U(x, t)$  and  $\tilde{U}(x, t)$  does not require to specify whether  $X$  and  $\tilde{X}$  are defined on the same probability space. On the other hand, the construction of bounds on the error  $|U(x, t) - \tilde{U}(x, t)|$  requires to specify probability spaces for  $X$  and  $\tilde{X}$ . Since the range  $(\tilde{x}^{(k)}, \dots, \tilde{x}^{(m)})$  of  $\tilde{X}$  consists of  $m$  samples of  $X$ , we view  $\tilde{X}$  as a measurable image of  $X$ .

We construct an approximate mapping  $X \mapsto \tilde{X} = h(X)$  under the assumption that  $X$  can be represented satisfactorily by a large number  $n$  of independent samples. In this setting, the construction of  $h$  is equivalent to that of a partition  $\{\mathcal{C}_k, k = 1, \dots, m\}$  of  $(x^{(1)}, \dots, x^{(n)})$  such that all  $x^{(i)}$  in  $\mathcal{C}_k$  are mapped into  $\tilde{x}^{(k)}$  and  $p_k \simeq n_k/n$  for all  $k = 1, \dots, m$ , where  $n_k$  is the cardinality of  $\mathcal{C}_k$ . The partition  $\{\mathcal{C}_k\}$  of  $(x^{(1)}, \dots, x^{(n)})$  can be constructed in two steps. First, construct a preliminary partition  $\mathcal{C}'_k$  of  $(x^{(1)}, \dots, x^{(n)})$  by assigning a sample  $x^{(i)}$  to  $\mathcal{C}'_k$  if it is closer to  $\tilde{x}^{(k)}$  than any other  $\tilde{x}^{(l)}$ ,  $l \neq k$ , that is,  $x^{(i)}$  is assigned to  $\mathcal{C}'_k$  if  $d(x^{(i)}, \tilde{x}^{(k)}) < d(x^{(i)}, \tilde{x}^{(l)})$ ,  $l \neq k$ , where  $d$  is a metric in the image space of  $X$ . If  $d(x^{(i)}, \tilde{x}^{(k)}) = d(x^{(i)}, \tilde{x}^{(l)})$  for a pair  $(k, l)$ ,  $x^{(i)}$  is assigned to either  $\mathcal{C}'_k$  or  $\mathcal{C}'_l$ . Generally, the cardinalities  $n'_k$  of  $\mathcal{C}'_k$  do not satisfy the condition  $p_k \simeq n'_k/n$ . Second, eliminate the members of the clusters  $\mathcal{C}'_k$  with  $n'_k/n > p_k$  that are the farthest from  $\tilde{x}^{(k)}$  until the reduced versions  $\mathcal{C}''_k$  of these clusters satisfy the condition  $n''_k/n \simeq p_k$ , where  $n''_k$  denotes the cardinality of  $\mathcal{C}''_k$ . The members extracted from the clusters  $\mathcal{C}'_k$  with  $n'_k/n > p_k$  are assigned to the clusters  $\mathcal{C}''_k$  with  $n'_k/n < p_k$  based on their closeness to the nuclei  $\tilde{x}^{(k)}$  of these clusters and the requirement  $n''_k/n \simeq p_k$ . The algorithm delivers a partition  $\{\mathcal{C}_k, k = 1, \dots, m\}$  of  $(x^{(1)}, \dots, x^{(n)})$  such that the members  $x^{(i)}$  of  $\mathcal{C}_k$  are mapped into  $\tilde{x}^{(k)}$  and the probability that  $X$  takes values in  $\mathcal{C}_k$  is  $n_k/n \simeq p_k$ . The resulting mapping is not unique. It depends on the particular sample  $(x^{(1)}, \dots, x^{(n)})$  used to characterize  $X$ , the model size  $m$ , and the metric use in  $S$ .

### 9.4.6.1 Initial-Boundary Value Problems

Suppose the spatial derivatives in (9.33) are approximated by finite differences, so that this equation becomes an ODE of the type in (9.27) with state vector  $Y(t)$  whose coordinates are values of  $U(x, t)$  at the nodes of the finite difference mesh. If the random coefficients in the definition of  $\mathcal{L}$  are time invariant,  $Y(t)$  is the solution of  $\dot{Y}(t) = A Y(t) + W(t)$  with initial state and driving noise  $W(t)$  inferred from  $U(x, 0)$  and  $V(x, t)$ , where  $A$  is a random matrix describing  $\mathcal{L}$ . Suppose  $X$  includes the random elements in the definition of  $\mathcal{L}$ . Let  $Y^{(i)}(t)$  and  $\tilde{Y}^{(k)}(t)$  be the solutions of  $\dot{Y}^{(i)}(t) = A^{(i)} Y^{(i)}(t) + W(t)$  and  $\dot{\tilde{Y}}^{(k)}(t) = \tilde{A}^{(k)} \tilde{Y}^{(k)}(t) + W(t)$ , where matrices  $A^{(i)}$  and  $\tilde{A}^{(k)}$  are equal to  $A$  with  $X$  replaced by  $x^{(i)}$  and  $\tilde{x}^{(k)}$ , respectively. The processes  $Y^{(i)}(t)$  and  $\tilde{Y}^{(k)}(t)$  are approximations for the solutions  $U^{(i)}(x, t)$  and  $\tilde{U}^{(k)}(x, t)$  of  $\mathcal{L}[U(x, t)] = V(x, t)$  with  $X$  set equal to  $x^{(i)}$  and  $\tilde{x}^{(k)}$ .

**Theorem 9.5** *Let  $Y(t)$  and  $\tilde{Y}(t)$  be the solutions of  $\dot{Y}(t) = A Y(t) + W(t)$  and  $\dot{\tilde{Y}}(t) = \tilde{A} \tilde{Y}(t) + W(t)$  as previously defined. Then*

$$\begin{aligned} E[ \| \tilde{Y}(t) - Y(t) \| ^q ] &\leq \frac{1}{n} \sum_{k=1}^m \sum_{x^{(i)} \in \mathcal{C}_k} \left( e^{\lambda_{i,\max} t} \int_0^t e^{-\lambda_{i,\max} s} \delta_{k,i}(s) ds \right)^q \\ &\simeq \sum_{k=1}^m p_k \left[ \frac{1}{n_k} \sum_{x^{(i)} \in \mathcal{C}_k} \left( e^{\lambda_{i,\max} t} \int_0^t e^{-\lambda_{i,\max} s} \delta_{k,i}(s) ds \right)^q \right], \end{aligned} \quad (9.70)$$

where  $\lambda_{i,\max}$  is the largest eigenvalue of  $(A^{(i)} + (A^{(i)})')/2$ ,  $\delta_{k,i}(t) = \| (A^{(i)} - \tilde{A}^{(k)}) \tilde{Y}^{(k)}(t) \|$ , and  $p_k \simeq n_k/n$ .

*Proof* Since  $U^{(i)}(x, t)$  and  $\tilde{U}^{(k)}(x, t)$  are driven by the same input, we have (7.115)

$$\| \tilde{Y}^{(k)}(t) - Y^{(i)}(t) \| \leq e^{\lambda_{i,\max} t} \int_0^t e^{-\lambda_{i,\max} s} \delta_{k,i}(s) ds, \quad k = 1, \dots, m, \quad i = 1, \dots, n. \quad (9.71)$$

If  $A^{(i)} = \tilde{A}^{(k)}$ , then  $\delta_{k,i}(s) = 0$  at all times, and so is the bound on  $\| \tilde{Y}^{(k)}(t) - Y^{(i)}(t) \|$ . If  $A^{(i)} \neq \tilde{A}^{(k)}$ , the bound on  $\| \tilde{Y}^{(k)}(t) - Y^{(i)}(t) \|$  can be calculated from (9.71). It depends on  $\tilde{Y}^{(k)}(t)$  over the time interval of interest, the real-valued function  $\delta_{k,i}(t)$ , and the largest eigenvalues  $\lambda_{i,\max}$  of  $(A^{(i)} + (A^{(i)})')/2$ , but does not depend on  $Y^{(i)}(t)$ ,  $i = 1, \dots, n$ . The inequalities in (9.71), the representation of  $X$  by  $n$  independent samples, and its SROM  $\tilde{X}$  yield the bound in (9.70).  $\blacktriangle$

Arguments yielding the bound in (9.70) can be extended directly to bound the discrepancy between the solutions  $Y(t)$  and  $\tilde{Y}(t)$  if  $X$  is characterized by its probability law rather than  $n$  independent samples. Let  $Y(t; a)$  denote the solution of  $\dot{Y}(t) = A Y(t) + W(t)$  with  $A$  set equal to  $A(x)$ ,  $x \in I = X(\Omega)$ , and  $\tilde{Y}^{(k)}(t)$  as previously. The bound in (9.71) gives

$$\| \tilde{Y}^{(k)}(t) - Y(t; x) \| \leq e^{\lambda_{\max}(x)t} \int_0^t e^{-\lambda_{\max}(x)s} \delta_k(s; x) ds, \quad (9.72)$$

where  $\lambda_{\max}(x)$  denotes the largest eigenvalue of  $(A(x) + A(x)')/2$  and  $\delta_k(s; x) = \| (A(x) - \tilde{A}^{(k)}) \tilde{Y}^{(k)}(t) \|$ . Let  $\{I_k, k = 1, \dots, m\}$  be a partition of the range  $I$  of  $X$  such that  $p_k \simeq \int_{I_k} f(x) dx$ , where  $f(x)$  denotes the density of  $X$ . Then, the expectation of  $\| \tilde{Y}(t) - Y(t) \|^q$  can be bounded by

$$\begin{aligned} E[ \| \tilde{Y}(t) - Y(t) \|^q ] &\leq E \left[ \left( e^{\lambda_{\max}(X)t} \int_0^t e^{-\lambda_{\max}(X)s} \delta_k(s; X) ds \right)^q \right] \\ &= \sum_{k=1}^m \int_{I_k} \left( e^{\lambda_{\max}(x)t} \int_0^t e^{-\lambda_{\max}(x)s} \delta_k(s; x) ds \right)^q f(x) dx. \end{aligned} \quad (9.73)$$

*Example 9.10* Let  $Y(t)$  be the solution of the stochastic ordinary differential equation  $\dot{Y}(t) = -A Y(t) + 1$ ,  $t \geq 0$ , where  $d = 1$ ,  $Y(0) = 0$ , and  $A$  is a shifted Gamma random variable with density  $f(z) = (z - \zeta)^{r-1} \xi^k e^{-\xi(z-\zeta)} / \Gamma(r)$ ,  $z - \zeta \geq 0$  depending on some constants  $\zeta, r, \xi > 0$ . The discrepancy between two deterministic solutions for  $A$  equal to  $a'$  and  $a''$  can be bounded by  $|Y(t; a') - Y(t; a'')| \leq e^{a't} \int_0^t e^{-a's} \delta(s; a', a'') ds$  where  $\delta(t; a', a'') = |a' - a''| |Y(t; a'')|$  by (9.72). Since the exact solution is  $Y(t) = (1 - \exp(-At))/A$ , then  $|Y(t; a') - Y(t; a'')|$  can be calculated exactly.

Let  $(\tilde{a}_1, \dots, \tilde{a}_m)$  and  $(p_1, \dots, p_m)$  be the defining parameters of a SROM  $\tilde{A}$  for  $A$ , and let  $(I_1, \dots, I_m)$  be a partition of the range  $[\zeta, \infty)$  of  $A$ ,  $\tilde{a}_k \in I_k$ , and  $p_k = P(A \in I_k)$ ,  $k = 1, \dots, m$ . Since  $\lambda_{\max}$  is equal to sample values of  $A$ ,  $\delta_{k,i}(t) = |a_i - \tilde{a}_k| |\tilde{Y}_k(t)|$  and  $\delta(t; a) = |a - \tilde{a}_k| |\tilde{Y}_k(t)|$ , the bound in (9.73) gives

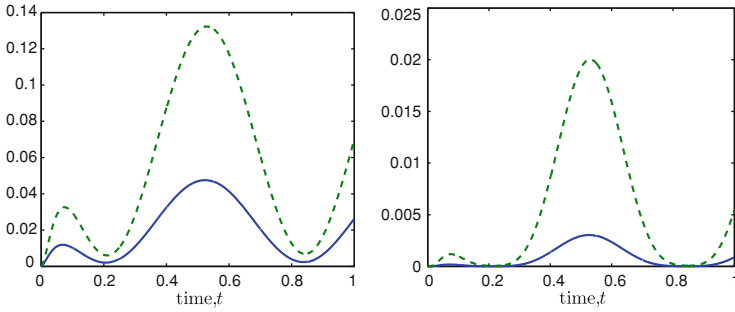
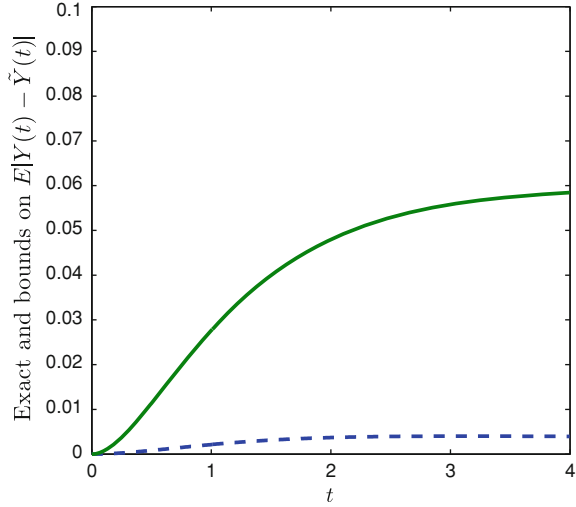
$$E[ |\tilde{Y}(t) - Y(t)|^q ] \leq \sum_{k=1}^m \int_{I_k} \left( e^{-a_i t} \left[ \int_0^t e^{a s} \delta_k(s; a) ds \right] \right)^q f_A(a) da. \quad (9.74)$$

Numerical results in Fig. 9.11 are for  $q = 1$ ,  $\zeta = 1$ ,  $r = 2$ ,  $\xi = 3$ ,  $m = 4$ ,  $I_1 = [0, 0.372)$ ,  $I_2 = [0.372, 1.020)$ ,  $I_3 = [1.020, 1.780)$ ,  $I_4 = [1.780, \infty)$ , and a SROM  $\tilde{A}$  of  $A$  with defining parameters  $(\tilde{a}_1, \dots, \tilde{a}_4) = (0.1640, 0.7253, 1.161, 2.4418)$  and  $(p_1, \dots, p_4) = (0.3068, 0.5028, 0.160, 0.0304)$ . The dotted and solid lines in Fig. 9.11 are the expectation  $E[|Y(t) - \tilde{Y}(t)|]$  and the bound on this expectation in (9.74) for  $\tilde{Y}(t)$  corresponding to a SROM  $\tilde{X}$  for  $X$  with  $m = 4$  samples.  $\diamond$

*Example 9.11* Let  $U(x, t)$ ,  $x \in (0, l)$ ,  $t \geq 0$ , be the solution of the SPDE in (9.60) with boundary conditions  $U(0, t) = U(l, t) = 0$  a.s. The coordinates of the vector  $Y(t)$  are values of  $U(x, t)$  at a finite number of coordinates  $x \in [0, l]$ . We have seen that  $Y(t)$  satisfies an ordinary differential equations of the type in (9.27).

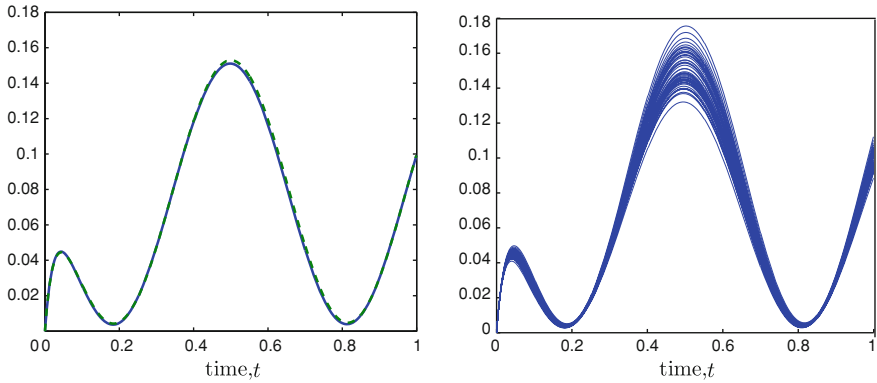
The following numerical results are for  $l = 1$ ,  $\lambda = 10$ , a uniform distribution  $F$  in the interval  $[a, b] = [2, 6]$ , and a partition of  $[0, l]$  in 11 equal intervals, so that  $Y(t)$  is an ten-dimensional vector if we account for the boundary conditions. Figure 9.12 shows

**Fig. 9.11** Exact expectation  $E[|Y(t) - \tilde{Y}(t)|]$  (dotted line) and the bound on this expectation in (9.74) (solid line)

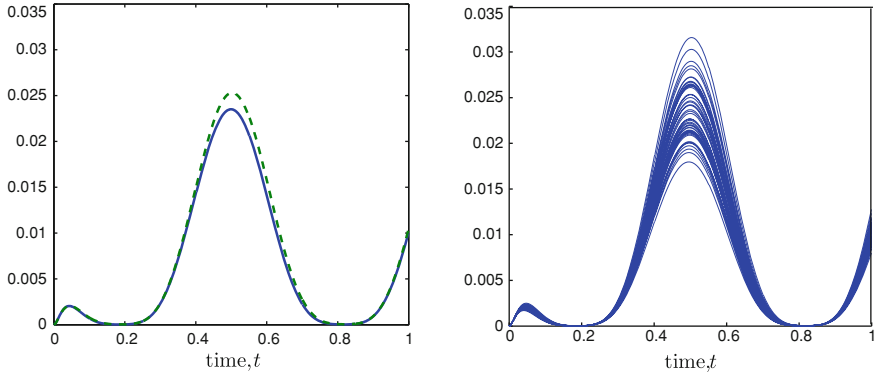


**Fig. 9.12** Monte Carlo estimates of  $E[\|\tilde{Y}(t) - Y(t)\|^q]$  (solid lines) and bounds on this expectation (dotted lines) for a SRM with  $m = 10$ ,  $q = 1$  (left panel), and  $q = 2$  (right panel)

with solid lines estimates of the expectation  $E[\|\tilde{Y}(t) - Y(t)\|^q]$  based on 500 samples of  $Y(t)$  corresponding to 500 independent samples of  $\Sigma(x)$  for  $q = 1$  (left panel) and  $q = 2$  (right panel) and a SRM  $\tilde{Y}(t)$  with  $m = 10$  samples for  $q = 1$  (left panel) and  $q = 2$  (right panel). The left panel in Fig. 9.13 shows with solid and dotted heavy lines an estimate of  $E[\|Y(t)\|]$  obtained from 500 independent samples of  $Y(t)$  and the expectation  $E[\|\tilde{Y}(t)\|]$  corresponding to a SRM of  $\Sigma(x)$  with  $m = 10$ . The thin solid lines in the right panel of the figure are estimates of  $E[\|Y(t)\|]$  calculated from 50 sets of ten independent samples of  $\Sigma(x)$ . In contrast to the SRM-based approximation of  $E[\|Y(t)\|]$ , approximations of this expectation based on independent samples of size ten exhibit significant sample-to-sample variation. Results similar to those in Fig. 9.13 are in Fig. 9.14 for the second moments of  $\|Y(t)\|$  and  $\|\tilde{Y}(t)\|$ . The estimates of  $E[\|Y(t)\|^2]$  in the right panel corresponding



**Fig. 9.13** Estimate of  $E[\|Y(t)\|]$  obtained from 500 samples (*solid heavy line, left panel*), an approximation of  $E[\|Y(t)\|]$  based on a SROM of  $\Sigma(x)$  with  $m = 10$  (*dotted heavy line, left panel*), and estimates of  $E[\|Y(t)\|]$  obtained from 50 sets of 10 independent samples of  $\Sigma(x)$  (*continuous thin lines, right panel*)



**Fig. 9.14** Estimate of  $E[\|Y(t)\|^2]$  obtained from 500 samples (*solid heavy line, left panel*), an approximation of  $E[\|Y(t)\|^2]$  based on a SROM of  $\Sigma(x)$  with  $m = 10$  (*dotted heavy line, left panel*), and estimates of  $E[\|Y(t)\|^2]$  obtained from 50 sets of 10 independent samples of  $\Sigma(x)$  (*continuous thin lines, right panel*)

to 50 sets of ten independent samples  $\Sigma(x)$  exhibit notable variation from one set to another.  $\diamond$

#### 9.4.6.2 Boundary Value Problems

Consider the stochastic boundary value problem in (9.53) and suppose that the random entries in this equation are collected in a random element  $X$  defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Let  $\tilde{X}$  be a SROM for  $X$  with defining parameters  $(\tilde{x}^{(k)}, p_k)$ ,  $k = 1, \dots, m$ . Suppose  $X$  can be characterized accurately by a large num-

ber  $n \gg m$  of independent, equally likely samples  $\{x^{(i)}\}$  and that these samples are clustered in subsets  $\mathcal{C}_k$  of  $(x^{(1)}, \dots, x^{(n)})$  with cardinality  $n_k$  such that  $p_k \simeq n_k/n$ .

The weak solutions  $U(x)$  and  $\tilde{U}(x)$  of (9.53) corresponding to the representation of  $X$  by  $n$  independent, equally likely samples  $\{x^{(i)}\}$  and its SROM  $\tilde{X}$  with parameters  $(\tilde{x}^{(k)}, p_k)$ ,  $k = 1, \dots, m$ , satisfy the equations  $\mathcal{B}(U, W) = \langle f, W \rangle_{L^2(D \times \Omega)}$  and  $\tilde{\mathcal{B}}(\tilde{U}, W) = \langle \tilde{f}, W \rangle_{L^2(D \times \Omega)}$ , where

$$\begin{aligned} \mathcal{B}(U, W) &= \frac{1}{n} \sum_{i=1}^n \int_D a^{(i)}(x) \nabla U^{(i)}(x) \cdot \nabla W^{(i)}(x) dx, \\ &= \sum_{k=1}^m p_k \left[ \frac{1}{n_k} \sum_{x^{(i)} \in \mathcal{C}_k} \int_D a^{(i)}(x) \nabla U^{(i)}(x) \cdot \nabla W^{(i)}(x) dx \right], \\ \tilde{\mathcal{B}}(\tilde{U}, W) &= \frac{1}{n} \sum_{k=1}^m p_k \left[ \frac{1}{n_k} \sum_{x^{(i)} \in \mathcal{C}_k} \int_D \tilde{a}^{(k)}(x) \nabla \tilde{U}^{(k)}(x) \cdot \nabla W^{(i)}(x) dx \right], \end{aligned} \quad (9.75)$$

$$\begin{aligned} \langle f, W \rangle_{L^2(D \times \Omega)} &= \sum_{k=1}^m p_k \left[ \frac{1}{n_k} \sum_{x^{(i)} \in \mathcal{C}_k} \int_D f^{(i)}(x) W^{(i)}(x) dx \right], \\ \langle \tilde{f}, W \rangle_{L^2(D \times \Omega)} &= \sum_{k=1}^m p_k \left[ \frac{1}{n_k} \sum_{x^{(i)} \in \mathcal{C}_k} \int_D \tilde{f}^{(k)}(x) W^{(i)}(x) dx \right], \end{aligned} \quad (9.76)$$

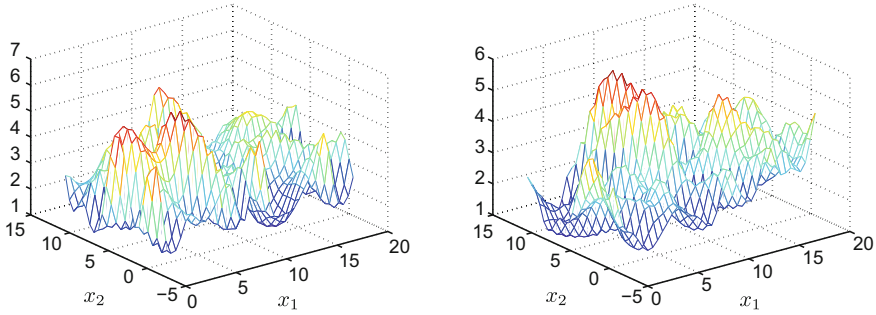
$U^{(i)}$ ,  $f^{(i)}$ ,  $a^{(i)}$ , and  $W^{(i)}$  correspond to  $x^{(i)}$ ,  $i = 1, \dots, n$ , and  $\tilde{U}^{(k)}$ ,  $\tilde{f}^{(k)}$ ,  $\tilde{a}^{(k)}$ , and  $\tilde{f}^{(k)}$  correspond to  $\tilde{x}^{(k)}$ ,  $k = 1, \dots, m$ .

The functionals in (9.75) and (9.76) satisfy the conditions of Theorem 9.4 so that the bound in (9.57) holds with

$$\begin{aligned} \|U - \tilde{U}\|_{\mathcal{H}(D, \Omega)}^2 &= \sum_{k=1}^m p_k \left[ \frac{1}{n_k} \sum_{x^{(i)} \in \mathcal{C}_k} \int_D (V^{(i,k)}(x)^2 + \nabla V^{(i,k)}(x) \cdot \nabla V^{(i,k)}(x)) dx \right] \\ \|a - \tilde{a}\|_{L^\infty(D \times \Omega)} &= \max_{1 \leq k \leq m} \max_{x^{(i)} \in \mathcal{C}_k} \sup_{x \in D} |a^{(i)}(x) - \tilde{a}^{(k)}(x)|, \\ \|\tilde{U}\|_{\mathcal{H}(D, \Omega)} &= \sum_{k=1}^m p_k \int_D (\tilde{U}^{(k)}(x)^2 + \nabla \tilde{U}^{(k)}(x) \cdot \nabla \tilde{U}^{(k)}(x)) dx, \quad \text{and} \\ \|f - \tilde{f}\|_{L^2(D \times \Omega)}^2 &= \sum_{k=1}^m p_k \left[ \frac{1}{n_k} \sum_{x^{(i)} \in \mathcal{C}_k} \int_D (f^{(i)}(x) - \tilde{f}^{(k)}(x))^2 dx \right], \end{aligned} \quad (9.77)$$

where  $V^{(i,k)}(x) = U^{(i)}(x) - \tilde{U}^{(k)}(x)$ , and  $f^{(i)}(x)$  and  $\tilde{f}^{(k)}(x)$  correspond to  $x^{(i)}$  and  $\tilde{x}^{(k)}$ , respectively.

**Example 9.12** Suppose the conductivity  $\Sigma(x)$  of a rectangular two-dimensional specimen  $D = (0, l_1) \times (0, l_2)$  can be modeled by a homogeneous random field. The



**Fig. 9.15** Two samples of a SROM  $\tilde{\Sigma}(x)$  with  $m = 10$  for  $\Sigma(x)$  with range  $[\alpha = 1, \beta = 8]$

potential  $U$  satisfies the stochastic boundary value problem  $-\nabla \cdot (\Sigma(x) \nabla U(x, \omega)) = 0$  in  $D$  with the boundary conditions  $U(0, x_2) = 0$ ,  $U(l_1, x_2) = 1$ ,  $\partial U(x_1, x_2)/\partial x_2 = 0$  for  $x_2 = 0$  and  $x_2 = l_2$ ,  $P$ -a.s. Our objective is to find properties of the effective conductivity  $\Sigma_{\text{eff}}$  approximately by SROMs. We have

$$\Sigma_{\text{eff}}(\omega) = \frac{1}{l_2} \int_D \Sigma(x, \omega) \frac{\partial U(x, \omega)}{\partial x_1} dx, \quad (9.78)$$

where  $U(x, \omega)$  denotes the potential in the specimen corresponding to a sample  $\Sigma(x, \omega)$  of  $\Sigma(x)$ .

It is assumed that  $\Sigma(x)$  is the Beta translation field in (9.62), so that it takes values in  $[\alpha, \beta]$ ,  $0 < \alpha < \beta < \infty$ , irrespective of the spectral density of its Gaussian image  $G(x)$ . The marginal moment of order  $r$  of  $\Sigma(x)$  is

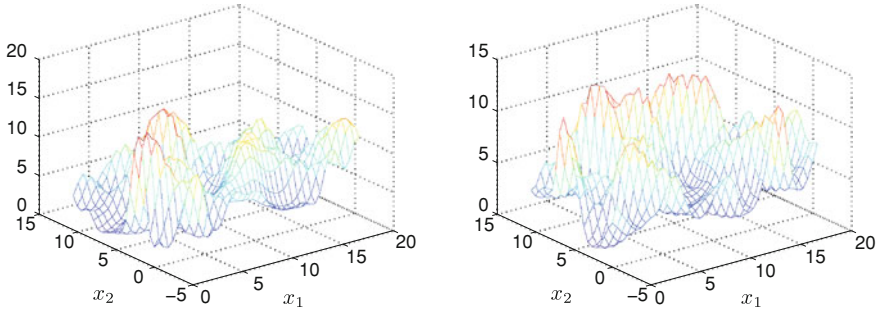
$$\mu(r) = E[\Sigma(x)^r] = \sum_{s=0}^r \frac{r!}{s!(r-s)!} (\beta - \alpha)^s \frac{B(p+s, q)}{B(p, q)} \alpha^{r-s}. \quad (9.79)$$

Two SROMs  $\tilde{\Sigma}(x)$  with  $m = 10$  and  $m = 20$  samples have been constructed for  $\Sigma(x)$  taking values in  $[\alpha = 1, \beta = 8]$  and  $[\alpha = 1, \beta = 20]$ , respectively. Consider first the case  $[\alpha = 1, \beta = 8]$  and  $m = 10$ . Figure 9.15 shows two samples of a SROM  $\tilde{\Sigma}(x)$  for  $\Sigma(x)$  recorded as sample  $\tilde{\sigma}_1(x)$  (left panel) and sample  $\tilde{\sigma}_3(x)$  (right panel) in our calculations. The probabilities of these samples of  $\tilde{\Sigma}(x)$  are  $p_1 = 0.1509$  and  $p_3 = 0.0701$ . The accuracy of  $\tilde{\Sigma}(x)$  is assessed by comparing spatial averages  $\tilde{\mu}(r) = \sum_{k=1}^m p_k (1/v_D) \int_D \tilde{\sigma}_k(x)^r dx$  of its marginal moments of order  $r$  with corresponding moments of  $\Sigma(x)$ . The first six moments of  $\tilde{\Sigma}(x)$  and  $\Sigma(x)$  are

$$\tilde{\mu}(r) = [0.0028 \ 0.0088 \ 0.0309 \ 0.1191 \ 0.4939 \ 2.1783] \times 10^3$$

$$\mu(r) = [0.0028 \ 0.0086 \ 0.0299 \ 0.1146 \ 0.4741 \ 2.0910] \times 10^3.$$

The largest error for the moments of  $\tilde{\Sigma}(x)$  based on a SROM  $\tilde{\Sigma}$  with  $m = 10$  samples is  $\max_{r=1, \dots, 6} \{100(\tilde{\mu}(r) - \mu(r))/\mu(r)\} = 4.18\%$ . Samples of  $\Sigma_{\text{eff}}$  and SROMs  $\tilde{\Sigma}_{\text{eff}}$



**Fig. 9.16** Two samples of a SROM  $\tilde{\Sigma}(x)$  with  $m = 20$  for  $\Sigma(x)$  with range  $[\alpha = 1, \beta = 20]$

of this random variable can be calculated from (9.78), samples  $\Sigma(x, \omega)$  of  $\Sigma(x)$ , and potentials  $U(x, \omega)$  corresponding to these samples. The first six moments  $\tilde{\mu}_{\text{eff}}(r)$  of the SROM-based approximation  $\tilde{\Sigma}_{\text{eff}}$  for  $\Sigma_{\text{eff}}$  and estimates  $\hat{\mu}_{\text{eff}}(r)$  of the first six moments  $E[\Sigma_{\text{eff}}^r]$  of  $\Sigma_{\text{eff}}$  obtained from  $n = 1000$  independent samples are

$$\begin{aligned}\tilde{\mu}_{\text{eff}}(r) &= [2.6924 \ 7.2749 \ 19.7232 \ 53.6437 \ 146.3420 \ 400.3602] \\ \hat{\mu}_{\text{eff}}(r) &= [2.6478 \ 7.0451 \ 18.8369 \ 50.6113 \ 136.6463 \ 370.7305]\end{aligned}$$

for  $r = 1, \dots, 6$ . The errors of  $\tilde{\mu}_{\text{eff}}(r)$  with respect to  $\hat{\mu}_{\text{eff}}(r)$  are 1.68%, 3.27%, 4.71%, 5.99%, 7.09%, and 7.99% for  $r = 1, 2, 3, 4, 5$ , and 6, respectively.

Suppose now that  $\Sigma(x)$  takes values in the range  $[\alpha = 1, \beta = 20]$ . Since  $\Sigma(x)$  has much larger variability in this case, we construct for  $\Sigma(x)$  a SROM  $\tilde{\Sigma}(x)$  with  $m = 20$ . Figure 9.16 shows two samples of  $\tilde{\Sigma}$  recorded in our calculations as sample  $\tilde{\sigma}_3(x)$  (left panel) and sample  $\tilde{\sigma}_{17}(x)$  (right panel). The probabilities of these samples of  $\tilde{\Sigma}$  are  $p_3 = 0.0662$  and  $p_{17} = 0.0390$ . The first six marginal moments  $\mu(r)$  of  $\Sigma(x)$  and of the spatial averages  $\tilde{\mu}(r)$  of the moments of  $\tilde{\Sigma}(x)$  are

$$\begin{aligned}\tilde{\mu}(r) &= [0.0001 \ 0.0004 \ 0.0033 \ 0.0309 \ 0.3124 \ 3.3913] \times 10^5 \\ \mu(r) &= [0.0001 \ 0.0004 \ 0.0035 \ 0.0325 \ 0.3283 \ 3.5515] \times 10^5\end{aligned}$$

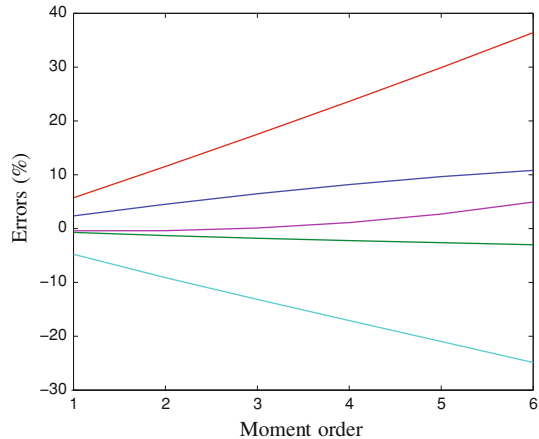
showing that the first six moments of  $\tilde{\Sigma}(x)$  are in error by  $-2.6298\%$ ,  $-4.1114\%$ ,  $-4.8654\%$ ,  $-5.1493\%$ ,  $-5.0822\%$ , and  $-4.7237\%$ . Since statistics of effective conductivity  $\Sigma_{\text{eff}}$  are not available analytically, they have been estimated from samples of this random variable. The first six moments  $\tilde{\mu}_{\text{eff}}(r)$ ,  $r = 1, \dots, 6$ , of  $\tilde{\Sigma}_{\text{eff}}$  and corresponding estimates  $\hat{\mu}_{\text{eff}}(r)$  of  $E[\Sigma_{\text{eff}}^r]$ ,  $r = 1, \dots, 6$ , obtained from 1000 independent samples are

$$\begin{aligned}\tilde{\mu}_{\text{eff}}(r) &= [0.0005 \ 0.0028 \ 0.0148 \ 0.0800 \ 0.4356 \ 2.3942] \times 10^4 \\ \hat{\mu}_{\text{eff}}(r) &= [0.0005 \ 0.0027 \ 0.0144 \ 0.0768 \ 0.4109 \ 2.2097] \times 10^4.\end{aligned}$$

The errors of  $\tilde{\mu}_{\text{eff}}(r)$  relative to  $\hat{\mu}_{\text{eff}}(r)$  are  $-0.60\%$ ,  $-1.45\%$ ,  $-2.57\%$ ,  $-3.98\%$ ,  $-5.69\%$ , and  $-5.69\%$ ,  $-7.70\%$  for  $r = 1, \dots, 6$ . Figure 9.17 shows estimates of



**Fig. 9.17** Estimates of first six moments of  $\Sigma_{\text{eff}}$  based on five sets of 20 samples each of this random variable



$E[\Sigma_{\text{eff}}^r]$ ,  $r = 1, \dots, 6$ , obtained from five sets of 20 independent samples of this random variable selected at random by Monte Carlo simulation. In contrast to SROM-based approximate moments for  $\Sigma_{\text{eff}}$ , Monte Carlo estimates can have very large errors and depend strongly on the particular set of 20 samples used for calculations.

The discrepancy between the SROM-based approximation  $\tilde{\Sigma}_{\text{eff}}$  and effective conductivity  $\Sigma_{\text{eff}}$  can be bounded by

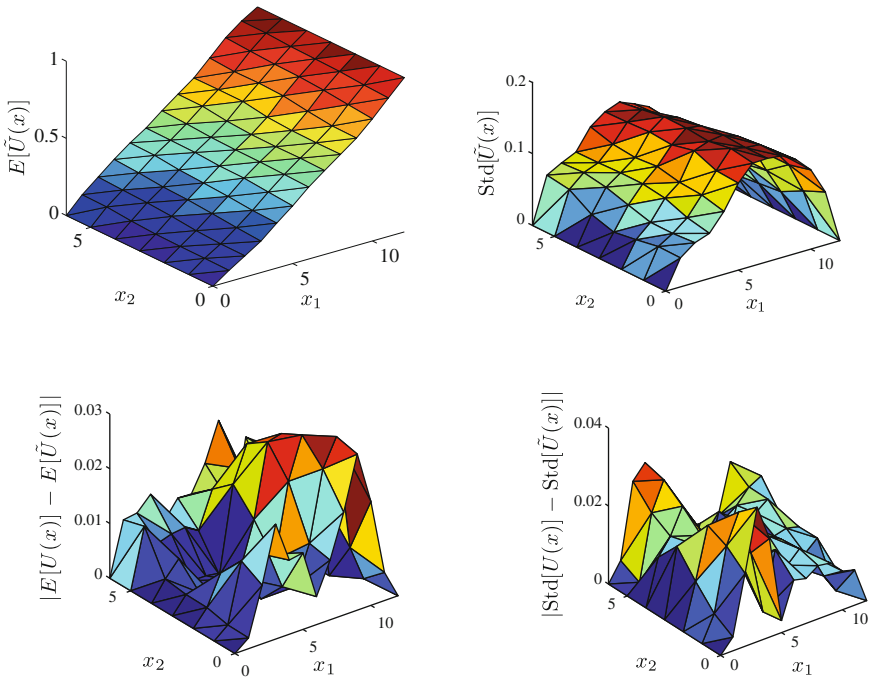
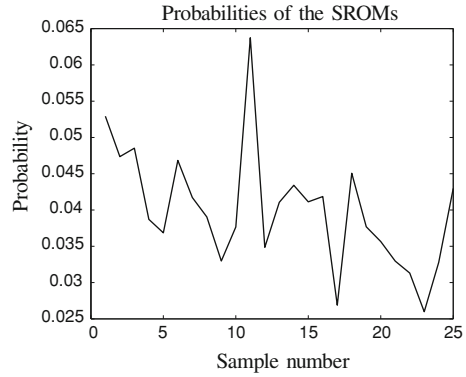
$$|\tilde{\Sigma}_{\text{eff}} - \Sigma_{\text{eff}}| \leq |\tilde{\Sigma}_{\text{eff}} - \hat{\Sigma}_{\text{eff}}| + |\hat{\Sigma}_{\text{eff}} - \Sigma_{\text{eff}}|, \quad (9.80)$$

where  $\hat{\Sigma}_{\text{eff}}$  denotes a Monte Carlo estimator for  $\Sigma_{\text{eff}}$  corresponding to  $n$  independent samples of  $\Sigma(x)$  and  $U(x)$ . Theorem 9.4 can be modified to obtain an upper bound on the error  $|\tilde{\Sigma}_{\text{eff}} - \hat{\Sigma}_{\text{eff}}|$  of the SROM-based approximation for  $\Sigma_{\text{eff}}$  with respect to Monte Carlo estimates based on  $n$  independent samples of  $\Sigma(x)$ . The second term on the right side of (9.80) can be made as small as desired by increasing  $n$ .  $\diamond$

**Example 9.13** Let  $U(x)$  be the solution of the stochastic partial differential equation (9.53) in Example 9.8 with  $a(x)$  in (9.62), (9.63), and (9.64). Let  $\tilde{a}(x)$  be a SROM for  $a(x)$  with samples  $(\tilde{a}_1(x), \dots, \tilde{a}_m(x))$  and probabilities  $(p_1, \dots, p_m)$ . Denote by  $(\tilde{u}_1(x), \dots, \tilde{u}_m(x))$  solutions of (9.53) with  $a(x)$  set equal to the samples  $(\tilde{a}_1(x), \dots, \tilde{a}_m(x))$  of  $\tilde{a}(x)$ , so that  $\tilde{U}(x)$  with samples  $(\tilde{u}_1(x), \dots, \tilde{u}_m(x))$  of probabilities  $(p_1, \dots, p_m)$  is a SROM for  $U(x)$ . Properties of  $\tilde{U}(x)$  can be obtained by elementary calculations. For example,  $E[\tilde{U}(x)^r] = \sum_{k=1}^m p_k \tilde{u}_k(x)^r$  is the moment of order  $r = 1, 2, \dots$  of  $\tilde{U}(x)$ .

Numerical results are for the parameter values in Example 9.8 and two SROMs for  $a(x)$ . The first SROM has  $m = 25$ . Figure 9.18 shows the probabilities  $\{p_k\}$  of these samples. The top panels in Fig. 9.19 show the spatial variation of the approximate means and standard deviations of  $\tilde{U}(x)$  corresponding to a SROM  $\tilde{a}(x)$  of  $a(x)$  with  $m = 25$  samples. The bottom panels in the figure gives absolute values of the errors of these moments relative to the Monte Carlo estimates of these moments in Example 9.8, that are based on 800 independent samples. The means and standard deviations

**Fig.9.18** Probabilities  $\{p_k\}$ ,  $k = 1, \dots, m$ , for a SROM of  $a(x)$  with  $m = 25$

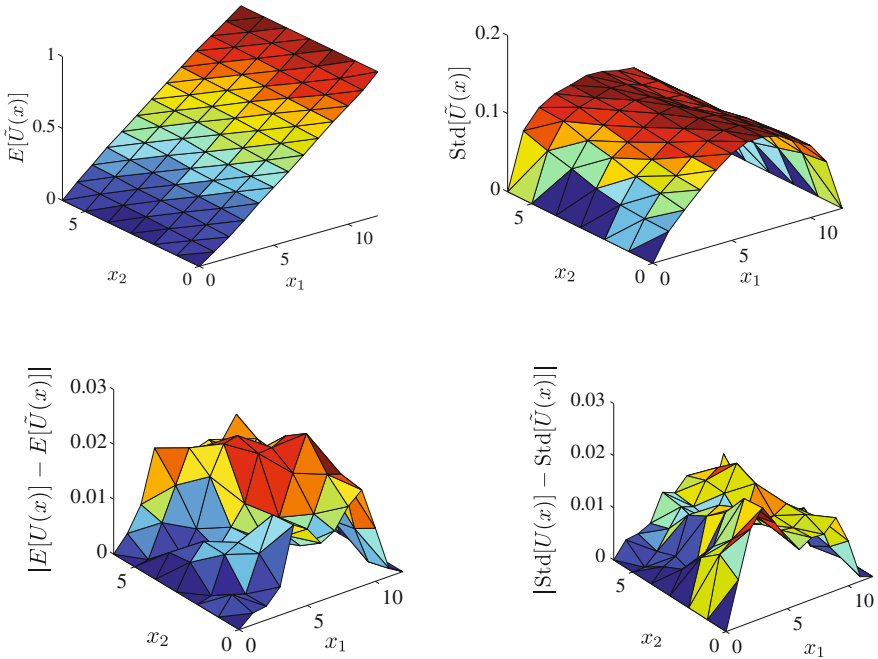
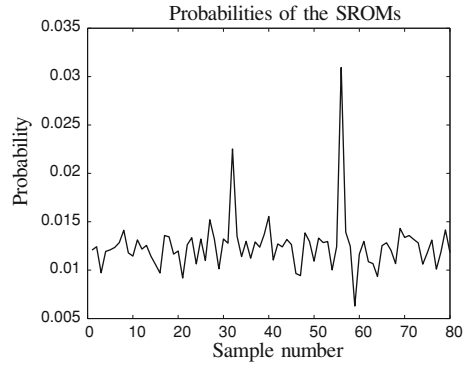


**Fig.9.19** Mean and standard deviations of  $\tilde{U}(x)$  for a SROM with  $m = 25$  samples (*top panels*) and errors relative to Monte Carlo estimates (*bottom panels*)

of the SROM  $\tilde{U}(x)$  are satisfactory, although they are based on  $m = 25$  samples, so that the construction of  $\tilde{U}(x)$  involved 25 solutions of deterministic versions of (9.53).

Similar plots are in Figs. 9.20 and 9.21, but they correspond to a SROM  $\tilde{a}(x)$  of  $a(x)$  with size  $m = 80$ . The probabilities  $\{p_k\}$  of the samples of  $\tilde{a}(x)$  are plotted in Fig. 9.20. The spatial variation of the mean and standard deviation of  $\tilde{U}(x)$  is

**Fig. 9.20** Probabilities  $\{p_k\}$ ,  $k = 1, \dots, m$ , for a SROM of  $a(x)$  with  $m = 80$



**Fig. 9.21** Mean and standard deviations of  $\tilde{U}(x)$  for a SROM with  $m = 80$  samples (*top panels*) and errors relative to Monte Carlo estimates (*bottom panels*)

shown in the top panels of Fig. 9.21. The absolute values of errors of the moments of  $\tilde{U}(x)$  relative to their Monte Carlo estimates, calculated in Example 9.8 and used in Fig. 9.19 as reference are in the bottom panels of the figure. The increase of model size from  $m = 25$  to  $m = 80$  results in superior SROM-based approximations. In addition to smaller errors, the spatial variation of the standard deviation of  $\tilde{U}(x)$  for  $m = 80$  traces that of the reference Monte Carlo estimates. The improvement of the performance of  $\tilde{U}(x)$  with the model size is rather slow, possibly because the

SROMs developed here are suboptimal. They have been extracted from  $n_{\text{set}} = 40$  sets of independent samples of  $a(x)$  of size  $m = 25$  and  $m = 80$ .  $\diamond$

### 9.4.7 Extended Stochastic Reduced Order Models

We have seen that the SROM-based method can be used to solve stochastic equations efficiently. The solution  $U$  of a stochastic equation is approximated by a SROM  $\tilde{U}$  with samples  $\{\tilde{u}_k\}$  corresponding to the samples of a SROM for the random elements in the definition of this equation. In Example 9.13, the samples  $\{\tilde{u}_k(x)\}$  of  $\tilde{U}(x)$  are solutions of deterministic PDEs with  $a(x)$  replaced by the samples  $\{\tilde{a}_k(x)\}$  of a SROM  $\tilde{a}(x)$  of it. A limitation of the approximate solution  $\tilde{U}(x)$  is that it does not explore the sensitivity of  $U(x)$  with respect to changes in  $\{\tilde{a}_k(x)\}$ . It is solely based on the expressions of  $U(x)$  corresponding to  $\{\tilde{a}_k(x)\}$ .

The extended version of the SROM-based method, referred to as the ESROM-based method, attempts to overcome this limitation. The method is discussed in Sect. A.4 for the case in which the random elements in the definition of a stochastic equation can be described by an  $\mathbb{R}^n$ -valued random variable  $C$  defined on a probability space  $(\Omega, \mathcal{F}, P)$ . The implementation of ESROM-based solutions involves the following three steps. First, a SROM  $\tilde{C}$  with samples  $\{\tilde{c}_k\}$ ,  $k = 1, \dots, m$ , is developed for  $C$ . Second, deterministic solvers are used to find solutions  $\{\tilde{u}_k\}$  corresponding to the samples  $\{\tilde{c}_k\}$  of  $\tilde{C}$  and gradients  $\{\nabla \tilde{u}_k\}$  of  $U$  with respect to the coordinates of  $C$  at  $\{\tilde{c}_k\}$ . The solution  $U$  is approximated by hyperplanes tangent to  $U$  at  $\{\tilde{c}_k\}$ , that is,

$$U_L(\cdot, C) = \sum_{k=1}^m [\tilde{u}_k(\cdot) + \nabla \tilde{u}_k(\cdot) \cdot (C - \tilde{c}_k)] 1(C \in \Gamma_k), \quad (9.81)$$

where  $\{\Gamma_k\}$  are the cells of a Voronoi tessellation with centers  $\{\tilde{c}_k\}$  constructed in the range  $\Gamma = C(\Omega)$  of  $C$ . If the mappings  $C \mapsto U, U_L$  are measurable, then  $U$  and  $U_L$  are random elements on  $(\Omega, \mathcal{F}, P)$ . Third, properties of  $U$  are approximated by those of  $U_L$  in (9.81), that can be estimated efficiently from samples of  $C$  since the functional form of  $U_L$  is available. The representation of  $U$  in (9.81) accounts not only for the expressions of  $U$  at  $\{\tilde{c}_k\}$ , as for SROM-based solutions, but also for the rate of change of  $U$  with the coordinates of  $C$  in vicinities of  $\{\tilde{c}_k\}$ .

The following example uses ESROMs to solve a version of the stochastic differential equation (9.53). Since  $a(x)$  in (9.53), as any other random field, consists of an uncountable number of random variables, it is represented approximately by a linear parametric model with dependent coefficients (Sect. 6.3.3.2).

*Example 9.14* Let  $U(x)$  be the solution of the stochastic partial differential equation (9.53) with  $a(x)$  in (9.62), (9.63), and (9.64). Let

$$\tilde{a}(x, C) = \sum_{i,j=0}^{n^*} A_{i,j} T_i(x_1) T_j(x_2) = \sum_{r=1}^n C_r \varphi_r(x), \quad x = (x_1, x_2) \in D, \quad (9.82)$$

be a linear parametric model with dependent coefficients (Sect. 6.3.3.2) for  $a(x)$ , where  $\{T_j(\cdot)\}$  are modified Chebyshev polynomials defined in  $[0,1]$  by the recurrence formula  $T_{j+1}(\xi) = 2(2\xi - 1)T_j(\xi) - T_{j-1}(\xi)$ ,  $j = 1, 2, \dots$ , with  $T_0(\xi) = 1$  and  $T_1(\xi) = 2\xi - 1$ , the functions  $\{\varphi_r(x)\}$  are products of Chebyshev polynomials in  $x_1$  and  $x_2$ , and  $C$  is an  $\mathbb{R}^n$ -valued random variable that is defined on a probability space  $(\Omega, \mathcal{F}, P)$  and collects the random coefficients  $\{A_{i,j}\}$ . Under the approximation  $a(x) \simeq \tilde{a}(x, C)$ , the stochastic differential equation (9.53) defines a mapping  $C \mapsto U$ . ESROM-based solutions constitute approximations for this mapping.

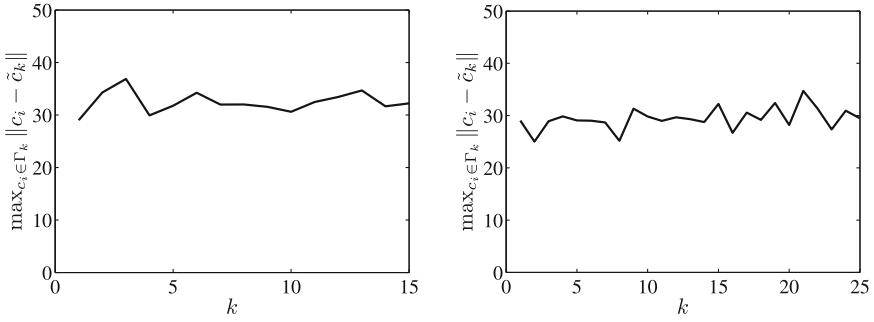
As previously mentioned, the first step of the ESROM method is the construction of SROMs for the uncertain parameters in the definition of a stochastic equation, that is, the random vector  $C$  for our case. Since the probability law of  $C$  is not known, samples of this vector are used to construct SROMs  $\tilde{C}$  for  $C$ . We calculate samples of  $C$  by minimizing the objective function in (6.48) that quantifies the discrepancy between  $a(x)$  and  $\tilde{a}(x, C)$ . The analysis is based on  $n_s = 1000$  independent samples of  $C$ .

Numerical results are reported for  $l_1 = 12$ ,  $l_2 = 6$ ,  $\alpha = 10$ ,  $\beta = 100$ ,  $p = 1$ ,  $q = 3$ ,  $\rho = 0.7$ , and stochastic dimension  $n = 25$ . Using an algorithm in [34],  $n_s = 1000$  independent, equally likely samples  $\{c_i\}$ ,  $i = 1, \dots, n_s$ , of  $C$  have been obtained by minimizing the discrepancy between samples of  $a(x)$  and  $\tilde{a}(x, C)$ . It is assumed that  $C$  is completely described by the samples  $\{c_i\}$ ,  $i = 1, \dots, n_s$ . This characterization of  $C$  is used to construct SROMs  $\tilde{C}$  for  $C$  and Monte Carlo estimates for properties of  $U$ . The construction of Monte Carlo estimates requires  $n_s = 1000$  solutions of distinct deterministic versions of (9.53) corresponding to  $C = c_i$ ,  $i = 1, \dots, n_s$ .

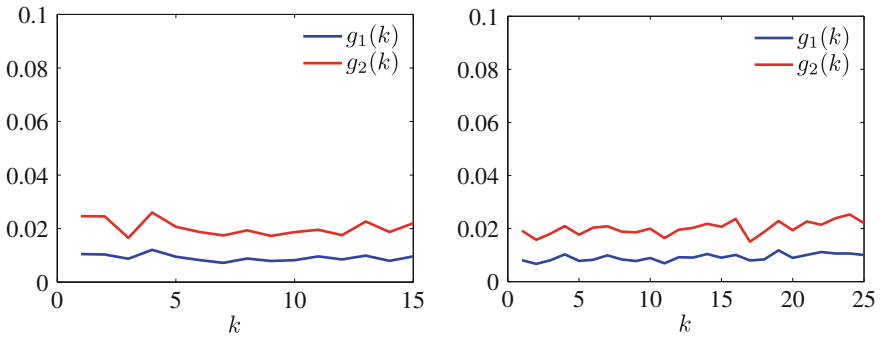
The construction of the piecewise linear representation  $U_L$  of  $U$  requires to find the deterministic functions  $\{\tilde{u}_k\}$  and  $\{\nabla \tilde{u}_k\}$ ,  $k = 1, \dots, m$ , that is, the functions  $U$  for  $m$  points  $\{\tilde{c}_k\}$  in the range  $\Gamma = C(\Omega)$  of  $C$  and the gradients of  $U$  with respect to the coordinates of  $C$  at  $\{\tilde{c}_k\}$ . The expansion points for  $\tilde{U}_L$  are the samples  $\{\tilde{c}_k\}$ ,  $k = 1, \dots, m$ , of a SROM  $\tilde{C}$  for  $C$  that are extracted from the set  $\{c_i\}$ ,  $i = 1, \dots, n_s$ , by an optimization algorithm [19]. The functions  $\{\tilde{u}_k\}$  are solutions of deterministic versions of (9.53) with  $\{\tilde{c}_k\}$  in place of  $C$ . The coordinates  $V_r(x, C) = \partial U(x, C) / \partial C_r$ ,  $r = 1, \dots, n$ , of the gradients of  $U$  at  $\{\tilde{c}_k\}$  are obtained from the differential equations

$$\nabla \cdot (\tilde{a}(x, C) \nabla V_r(x, C)) = -\nabla \cdot \left( \frac{\tilde{a}(x, C)}{\partial C_r} \nabla U(x, C) \right), \quad r = 1, \dots, n, \quad (9.83)$$

with  $\tilde{c}_k$ ,  $k = 1, \dots, m$ , in place of  $C$ . The latter equation is derived from (9.53) by differentiation with respect to the coordinates  $C_r$ ,  $r = 1, \dots, n$ , of  $C$ . Since (9.83) and (9.53) have the same differential operator, their solutions for  $n + 1$  distinct right sides involves a single inversion of their common operator, so that only  $m$  distinct deterministic operators need to be inverted. This observation increases significantly



**Fig. 9.22** An approximate measure for cell size corresponding to SROMs with  $m = 15$  (left panel) and  $m = 25$  (right panel)



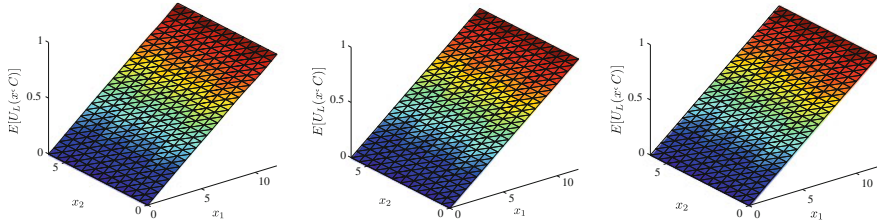
**Fig. 9.23** Measures  $g_1(k)$  and  $g_2(k)$  of local gradients  $\nabla \tilde{u}_x(x)$  for  $m = 15$  (left panel) and  $m = 25$  (right panel)

the efficiency of the finite element algorithm used for calculations. Since the mapping  $C \mapsto U_L$  is known and has a simple functional form, statistics of  $\tilde{U}_L$  can be obtained from samples of  $C$  with a minimum computational effort. The Euclidean distance in  $\mathbb{R}^n$  is used to assign the samples  $\{c_i\}$  of  $C$  to the Voronoi cells  $\{\Gamma_k\}$ . For example,  $c_i$  is assigned to  $\Gamma_k$  if  $\|c_i - \tilde{c}_k\| < \|c_i - \tilde{c}_l\|$  for all  $l \neq k$ . If  $\|c_i - \tilde{c}_k\| = \|c_i - \tilde{c}_l\|$ ,  $c_i$  can be assigned to either  $\Gamma_k$  or  $\Gamma_l$ . Note that it is not necessary to construct Voronoi tessellations explicitly for the range  $\Gamma = C(\Omega)$  of  $C$  to implement ESROM-based solutions.

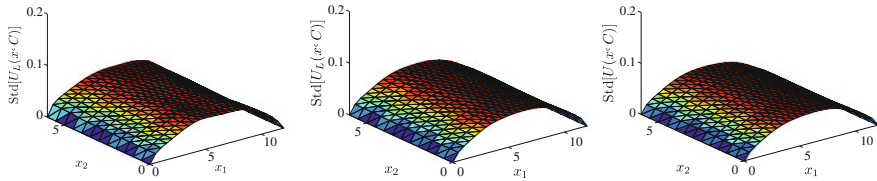
The approximate measure  $\max_{c_i \in \Gamma_k} \|c_i - \tilde{c}_k\|$  for the size of the Voronoi cells is shown in the left and right panels of Fig. 9.22 for SROMs with  $m = 15$  and  $m = 25$ , respectively. As expected, the cells for  $m = 25$  are smaller on average than those for  $m = 15$ . The magnitude of the gradients of  $U(x, C)$  at  $\{\tilde{c}_k\}$  is quantified by

$$g_1(k) = \max_{1 \leq r \leq n} \max_{x \in D} |\partial \tilde{u}_k(x) / \partial C_r| \text{ and}$$

$$g_2(k) = \left( (1/\text{vol}(D)) \int_D \|\nabla \tilde{u}_k(x)\|^2 dx \right)^{1/2}$$



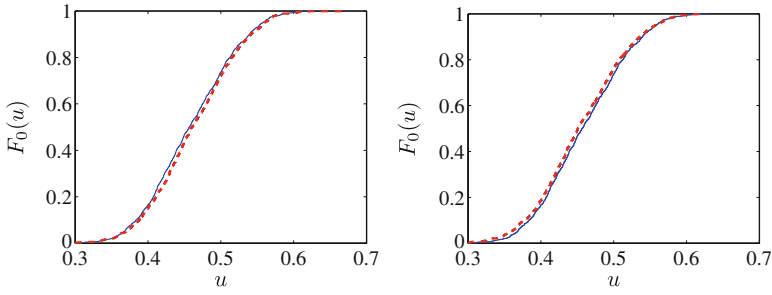
**Fig. 9.24** Approximations of  $E[U(x, C)]$  based on a SROM with  $m = 15$  (left panel), a SROM with  $m = 25$  (middle panel), and Monte Carlo simulation (right panel)



**Fig. 9.25** Approximations of  $\text{Std}[U(x, C)]$  based on a SROM with  $m = 15$  (left panel), a SROM with  $m = 25$  (middle panel), and Monte Carlo simulation (right panel)

and shown in Fig. 9.23 for  $m = 15$  (left panel) and  $m = 25$  (right panel). Since the cells  $\{\Gamma_k\}$  have similar sizes and the gradients  $\{\nabla \tilde{u}_k\}$  have similar magnitudes, we conclude that there is no need to refine the partition  $\{\Gamma_k\}$  of  $\Gamma$ .

Figure 9.24 shows approximations for  $E[U(x, C)]$  based on a SROM with  $m = 15$  (left panel), a SROM with  $m = 25$  (middle panel), and Monte Carlo simulation (right panel). The Monte Carlo estimates are based on 1000 samples of  $U(x, C)$ . Standard deviations of  $U(x, C)$  are in Fig. 9.25. The plots in the left, middle, and right panels are based on a SROM with  $m = 15$ , a SROM with  $m = 25$ , and Monte Carlo simulation using 1000 samples of  $U(x, C)$ . The largest discrepancy between the means of  $U(x, C)$  by SROMs relative to Monte Carlo estimates are 0.012 for  $m = 15$  and  $5.5 \times 10^{-3}$  for  $m = 25$ . Corresponding discrepancies for standard deviations are 0.095 for  $m = 15$  and  $4.8 \times 10^{-3}$  for  $m = 25$ . The increase of the model size from  $m = 15$  to  $m = 25$  improves significantly the accuracy of the SROM-based approximations for  $E[U(x, C)]$  and  $\text{Std}[U(x, C)]$ . The thin solid lines in Fig. 9.26 coincide and give a Monte Carlo estimate for the distribution  $F_0(u) = P(U(x_0, C) \leq u)$  of  $U(x, C)$  at  $x_0 = (l_1/2, l_2/2)$  obtained from  $n_s = 1000$  independent samples of  $U(x_0)$ . The heavy dotted lines in the figures are approximations of  $F_0(u)$  corresponding to ESROM-based solutions with  $m = 15$  (left panel) and  $m = 25$  (right panel).  $\diamond$



**Fig. 9.26** A Monte Carlo estimate for  $F_0(u) = P(U(x_0, C) \leq u)$ ,  $x_0 = (l_1/2, l_2/2)$ , (think solid lines) and approximations of  $F_0(u)$  based on SROMs with  $m = 15$  (left panel) and  $m = 25$  (right panel)

### 9.4.8 Stochastic Galerkin Method

The Galerkin method has been used extensively in applications to solve a broad range of stochastic problems in physics and engineering [25]. We review essentials of the Galerkin method and evaluate its performance. The stochastic elliptic boundary value problem (9.53) is the model problem considered in our discussion.

Let  $a(x)$  and  $f(x)$ ,  $x \in D$ , in (9.53) be real-valued random fields defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Under some conditions stated in Sect. 9.4.8.3,  $a(x)$  and  $f(x)$  admit Karhunen-Loève (KL) representations and (9.53) has a unique weak solution. The weak solution  $U(x, \omega)$  of this equation satisfies (9.54) and belongs to  $\mathcal{W}(D, \Omega)$  defined in (9.55). Generally, numerical methods need to be employed to solve (9.53), since this equation can be solved analytically only in special cases.

The implementation of numerical methods requires the discretization of both the probability space and the physical space. Methods for discretizing these spaces are discussed in the following two sections. Subsequent sections deal with the existence and uniqueness of Galerkin solutions and the construction of parametric models for the random elements in (9.53). Numerical examples conclude our discussion on the stochastic Galerkin method.

#### 9.4.8.1 Probability Space Discretization

The discretization of probability space involves two steps. First, the random fields in the definition of a stochastic problem are approximated by parametric models, that is, deterministic functions of space and/or time arguments depending on a finite number of random variables. Second, the random variables in the definition of a parametric model are approximated by their projections on subspaces of the space random variables with finite variance.

*First step.* Random functions, that is, uncountable families of random variables, are approximated by parametric random functions. We consider parametric models given by truncated KL representations and by parametric translation functions. Para-



metric models can also be constructed from a sampling theorem for random functions whose spectral densities have bounded support ([26] and Sect. A.1.3 in this book).

Consider first KL-based parametric models. Suppose that the random fields  $a(x)$  and  $f(x)$  in (9.53) are defined on a probability space  $(\Omega, \mathcal{F}, P)$  and that they admit KL representations. Truncated versions of these representations have the form

$$\begin{aligned} a(x) &\simeq a(x, Z) = \sum_{k=1}^m a_k(x) Z_k \\ f(x) &\simeq f(x, Z) = \sum_{k=1}^m f_k(x) Z_k, \quad x \in D, \end{aligned} \quad (9.84)$$

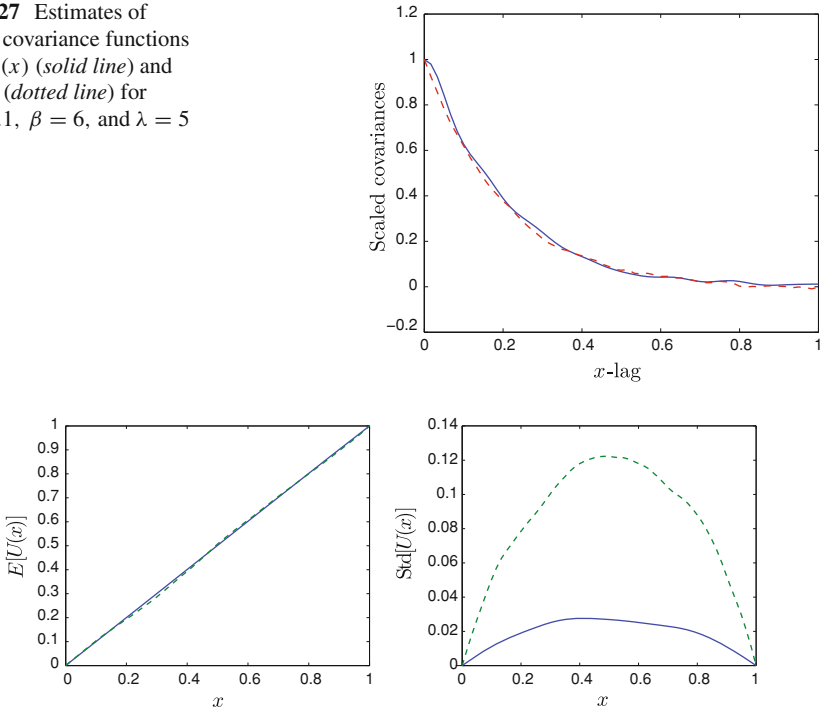
where  $Z = (Z_1, \dots, Z_m)$  is an  $\mathbb{R}^m$ -valued random variable defined on  $(\Omega, \mathcal{F}, P)$  with uncorrelated but, generally, dependent coordinates. The deterministic functions  $\{a_k(x)\}$  and  $\{f_k(x)\}$  result from the KL expansions of these fields. We refer to (9.84) as linear parametric models since they are finite sums of known deterministic functions of  $x \in D$  with random coefficients. If  $a(x)$  and  $f(x)$  are dependent, their KL expansion needs to be constructed jointly. If  $a(x)$  and  $f(x)$  are independent, their parametric models will depend on disjoint sets of coordinates of  $Z$  that are independent. For simplicity, we use the formulas in (9.84) irrespective of whether  $a(x)$  and  $f(x)$  are or are not independent.

The parametric models  $a(x, Z)$  and  $f(x, Z)$  in (9.84) are essential for the solution of stochastic problems since they depend on  $m < \infty$  of random variables. The KL-based linear parametric models are partially specified, in the sense that only the first two moments of  $Z$  and, therefore,  $a(x, Z)$  and  $f(x, Z)$  are known. This characterization is insufficient to generate samples of  $a(x, Z)$  and  $f(x, Z)$  and establish conditions for the existence and uniqueness for solutions of stochastic problem. It is common to augment the partial characterization of  $Z$  by assuming that its coordinates are independent and follow specified distributions such that (9.53) with  $a(x, Z)$  in place of  $a(x)$  is elliptic almost surely [16, 20, 25, 27–29, 30–33]. The selection of the probability law for  $Z$  based solely on mathematical considerations results in parametric models for  $a(x)$  and  $f(x)$  that may be of limited practical value (Sect. 6.3.3.1). We note that linear parametric models for  $a(x)$  and  $f(x)$  can be constructed such that they satisfy both mathematical and physical constraints ([34] and Sect. 6.3.3.2 in this book).

The following example shows an additional limitation of KL representations. It relates to the facts that random functions with identical second moment properties may have very different sample properties. These differences may be amplified by the solutions of SPDEs since they constitute nonlinear mappings of their random coefficients.

*Example 9.15* The solution of the SPDE  $d(a(x)U'(x)) = 0$ ,  $x \in (0, l)$ , with boundary conditions  $U(0) = 0$  and  $U(1) = 1$  is  $U(x) = \int_0^x K(u) du / \int_0^l K(u) du$ , where  $a(x) > 0$  a.s. denotes a random field and  $K(x) = 1/a(x)$  is assumed to have finite variance. We consider two models for  $K(x)$ , the random fields  $K_1(x)$  and  $K_2(x)$ .

**Fig. 9.27** Estimates of scaled covariance functions for  $K_1(x)$  (solid line) and  $K_2(x)$  (dotted line) for  $\alpha = 0.1$ ,  $\beta = 6$ , and  $\lambda = 5$



**Fig. 9.28** Mean and standard deviations of  $U(x)$  for  $K(x) = K_1(x)$  (solid lines) and  $K(x) = K_2(x)$  (dotted lines) for  $a = 0.1$ ,  $b = 6$ , and  $\lambda = 5$

The first model is the translation field  $K_1(x) = \alpha + (\beta - \alpha)\Phi(G(x))$ , where  $0 < \alpha < \beta < \infty$  and  $G(x)$  is a homogeneous Gaussian field with mean 0, variance 1, and covariance function  $E[G(x)G(y)] = \exp(-\lambda|x - y|)$ ,  $\lambda > 0$ . The mean and variance of  $K_1(x)$  are  $E[K_1(x)] = (\alpha + \beta)/2$  and  $\text{Var}[K_1(x)] = (\beta - \alpha)^2/12$ . The covariance function of  $K_1(x)$  can be calculated from its definition and the probability law of  $G(x)$ . The covariance of  $K_1(x)$  scaled by its variance is approximately equal to that of  $G(x)$ , and we assume it coincides with the covariance function of  $G(x)$ .

The second model is a homogeneous random field  $K_2(x)$  whose samples are piecewise constant with jumps at Poisson points spaced on average at  $1/\lambda$ . The marginal density of  $K_2(x)$  is  $f_2(z) = 1(\mu - \eta - \varepsilon < z < \mu - \eta + \varepsilon)/(4\varepsilon) + 1(\mu + \eta - \varepsilon < z < \mu + \eta + \varepsilon)/(4\varepsilon)$ ,  $0 < \varepsilon < \eta$ , so that its mean and variance are  $(\alpha + \beta)/2$  and  $\varepsilon^2/3 + \eta^2$ . Under the conditions  $\text{Var}[K_1(x)] = \text{Var}[K_2(x)]$  and  $\eta^2 = 0.99\text{Var}[K_1(x)]$ , the random fields  $K_1(x)$  and  $K_2(x)$  have the same second moment properties (Exercise 9.13), so that they admit the same KL representation.

Numerical results in the following figures are for  $\alpha = 0.1$ ,  $\beta = 6$ , and  $\lambda = 5$ . Figure 9.27 shows with solid and dotted lines estimates of the scaled covariance functions of  $K_1(x)$  and  $K_2(x)$ . Figure 9.28 shows with solid and dotted lines the

means and standard deviations of the solutions  $U_i(x)$  corresponding to  $K(x) = K_i(x)$ ,  $i = 1, 2$ . The two solutions have similar expectations but their standard deviations differ significantly. In summary, the random fields  $K_1(x)$  and  $K_2(x)$  are equal in the second moment sense, so that they admit the same KL representation implying that the resulting stochastic Galerkin solutions for these random coefficients coincide. Yet, the standard deviations of  $U_1(x)$  and  $U_2(x)$  corresponding to these random coefficients differ significantly.  $\diamond$

Consider now parametric translation models. Suppose that  $a(x)$  in (9.53) is a weakly homogeneous random field specified partially by its marginal distribution  $F$  and correlation function  $r_a(y) = E[a(x+y)a(x)]$ . The homogeneity assumption can be relaxed. Let

$$a(x) \simeq a_T(x) = F^{-1} \circ \Phi(G(x)) = h(G(x)), \quad x \in D, \quad (9.85)$$

be a translation field, where  $G(x)$  is a stationary Gaussian field with mean 0, variance 1, and covariance function  $\rho(y) = E[G(x+y)G(x)]$ . The correlation function of  $a_T(x)$  is

$$\begin{aligned} r_{a_T}(y) &= E[a_T(x+y)a_T(x)] = E[h(G(x+y))h(G(x))] \\ &= \int_{\mathbb{R}^2} h(u)h(v)\phi(u, v; \rho(y)) du dv, \end{aligned} \quad (9.86)$$

and depends on  $\rho(y)$  and the mapping in (9.85), where  $\phi(\cdot, \cdot; \lambda)$  denotes the joint density of two correlated  $N(0,1)$  variables with correlation coefficient  $\lambda$ . The marginal distribution of  $a_T(x)$  is  $F$  irrespective of the covariance function of  $G(x)$ , so that, if the support of  $F$  is a bounded interval  $[\alpha, \beta]$ ,  $0 < \alpha \leq \beta < \infty$ , the samples of  $a_T(x)$  are in  $[\alpha, \beta]$  with probability 1. On the other hand, there may not exist  $\rho(y)$  such that  $r_{a_T}(y) = r_a(y)$  ([22], Sect. 3.1). Optimization algorithms can be used to select  $\rho(y)$  such that the discrepancy between  $r_{a_T}(y)$  and  $r_a(y)$  is minimize in some sense (Sect. 3.8.2).

Parametric translation models for  $a(t)$  are given by its parametric translation model in (9.85) with  $G^{(n)}(x)$  in place of  $G(x)$ , that is, the random field

$$a_T^{(n)}(x) = F^{-1} \circ \Phi(G^{(n)}(x)) = h(G^{(n)}(x)), \quad x \in D, \quad (9.87)$$

where  $G^{(n)}(x)$ ,  $n = 1, 2, \dots$ , is a sequence of homogeneous parametric Gaussian fields with mean  $E[G^{(n)}(x)] = 0$ , variance  $E[G^{(n)}(x)^2] = 1$ , and covariance function  $\rho^{(n)}(y) = E[G^{(n)}(x+y)G^{(n)}(x)]$ . Discrete spectral representations or truncated KL expansions (Sects. 3.6.4 and 3.6.5) can be used to construct the fields  $\{G^{(n)}(x)\}$  such that the covariance function of  $G^{(n)}(x)$  converges to that of  $G(x)$  as  $n \rightarrow \infty$ .

**Theorem 9.6** *The parametric translation model  $a_T^{(n)}(x)$  in (9.87) is a homogeneous field with marginal distribution  $F$  that becomes a version of  $a_T(x)$  as  $n \rightarrow \infty$ . If  $G^{(n)}(x)$  has continuous samples and the mapping  $G(x) \mapsto a_T(x) = h(G(x))$  in (9.87) is continuous, then  $a_T^{(n)}(x)$  has a.s. continuous samples.*

*Proof* The finite dimensional distributions of  $a_T^{(n)}(x)$  are  $P(\cap_{i=1}^q a_T^{(n)}(x_i) \leq z_i) = P(\cap_{i=1}^q G^{(n)}(x_i) \leq \Phi^{-1} \circ F(z_i))$ , where  $q \geq 1$  is an arbitrary integer and  $z_i \in \mathbb{R}$ . Since  $G^{(n)}(x)$  is weakly homogeneous and Gaussian, it is a homogeneous Gaussian field, and so is  $a_T^{(n)}(x)$ .

By the normal comparison lemma ([35], Theorem 4.2.1) we have

$$\begin{aligned} & |P(\cap_{i=1}^q a_T^{(n)}(x_i) \leq z_i) - P(\cap_{i=1}^q a_T(x_i) \leq z_i)| \\ &= |P(\cap_{i=1}^q G^{(n)}(x_i) \leq \xi_i) - P(\cap_{i=1}^q G(x_i) \leq \xi_i)| \\ &\leq \frac{1}{2\pi} \sum_{1 \leq i < j \leq q} |\rho^{(n)}(x_i - x_j) - \rho(x_i - x_j)| (1 - \alpha_{ij}^2)^{-1/2} \exp\left(-\frac{(\xi_i^2 + \xi_j^2)/2}{1 + \alpha_{ij}}\right), \end{aligned}$$

where  $\alpha_{ij} = \max(|\rho^{(n)}(x_i - x_j)|, |\rho(x_i - x_j)|)$  and  $\xi_i = \Phi^{-1} \circ F(z_i)$ . The convergence of the covariance function of  $G^{(n)}(x)$  to that of  $G(X)$  as  $n \rightarrow \infty$  implies  $P(\cap_{i=1}^q a_T^{(n)}(x_i) \leq z_i) \rightarrow P(\cap_{i=1}^q a_T(x_i) \leq z_i)$ , that is, the finite dimensional distributions of  $a_T^{(n)}(x)$  converge to those of  $a_T(x)$ , so that  $a_T^{(n)}(x)$  becomes a version of  $a_T(x)$  as  $n \rightarrow \infty$ . If the samples  $G^{(n)}(x, \omega)$  of  $G^{(n)}(x)$  are continuous functions for  $\omega \in \Omega \setminus \Omega_0$ ,  $P(\Omega_0) = 0$ , then  $a_T^{(n)}(x, \omega) = h(G^{(n)}(x, \omega))$  are continuous functions by the postulated continuity of the mapping in (9.87).  $\blacktriangle$

*Second step.* We have seen that parametric models of the type in (9.84) can be constructed for the random fields  $a(x)$  and  $f(x)$  in the definition of the stochastic boundary value problem defined by (9.53). Also, it is possible to construct a measurable mapping  $g : \mathbb{R}^m \rightarrow \Gamma \subset \mathbb{R}^m$  relating  $Z$  to an  $m$ -dimensional random vector  $G = (G_1, \dots, G_m)$  defined on the same probability space  $(\Omega, \mathcal{F}, P)$  since  $Z = g(G)$  by (8.30). The coordinates of  $Z$  can be approximated by

$$Z_k = g_k(G) \simeq \sum_{i=0}^{n_{\text{PC}}} \alpha_{k,i} \psi_i(G), \quad k = 1, \dots, m \quad (9.88)$$

where  $\{\alpha_{k,i}\}$  are coefficients,  $\{\psi_i(G)\}$  denote polynomial chaoses (Sect. B.6.2), and  $n_{\text{PC}}$  is the largest degree of polynomial chaoses in the representation of  $Z_k$ . This representation and (9.84) give

$$\begin{aligned} a(x) &\simeq a(x, Z) \simeq \sum_{i=0}^{n_{\text{PC}}} \left[ \sum_{k=1}^m \alpha_{k,i} a_k(x) \right] \psi_i(G) \\ f(x) &\simeq f(x, Z) \simeq \sum_{i=0}^{n_{\text{PC}}} \left[ \sum_{k=1}^m \alpha_{k,i} f_k(x) \right] \psi_i(G), \quad x \in D. \end{aligned} \quad (9.89)$$

*Example 9.16* Let  $Z \sim U(0, 1)$  and  $G \sim N(0, 1)$  so that  $Z = g(G) = \Phi(G)$ . Let  $Z \simeq Z^{(n_{\text{PC}})} = \sum_{k=1}^{n_{\text{PC}}} \alpha_k H_k(G)$ , where  $\alpha_k$  and  $H_k$  are coefficients and Hermite polynomials (Sect. B.6.1). Since  $H_k$  are polynomials and  $G$  takes values over the entire real line, the support of  $Z^{(n_{\text{PC}})}$  is also the real line, so that the distributions of

$Z^{(n_{PC})}$  and  $Z$  differ significantly. The probability law of  $G$  and the mapping  $G \mapsto Z = \Phi(G)$  are inconsistent with the distribution  $Z \sim U(0, 1)$  postulated for  $Z$ . Alternative polynomials and distributions need to be considered (Example 8.14).  $\diamond$

Note that the parametric translation models (9.87) can also be used to construct polynomial chaos representations for the random functions in the definition of stochastic equations similar to those in (9.88), since they are nonlinear functions of  $x \in D$  depending on a finite number of independent Gaussian variables.

#### 9.4.8.2 Physical Space Discretization

The weak solution  $U(x)$  of (9.53) belongs to  $\mathcal{W}(D, \Omega)$ , so that  $U$  and  $\nabla U$  are in  $L^2(D \times \Omega)$ . Since  $U(\cdot, \omega)$  has an infinite number of degrees of freedom, we construct approximations for this function in a subspace of  $L^2(D)$  spanned by a finite family of functions  $(\zeta_1(x), \dots, \zeta_{n_e}(x))$ , where  $\zeta_j : D \rightarrow \mathbb{R}$ ,  $j = 1, \dots, n_e$ , are square integrable in  $D$ .

The numerical solution of (9.53) is sought in a finite dimensional subspace of  $\mathcal{W}(D, \Omega)$  spanned by  $\{\psi_i(G)\zeta_j(x)\}$ ,  $i = 0, 1, \dots, n_{PC}$ ,  $j = 1, \dots, n_e$ , so that the projection of  $U(x, \omega)$  on this subspace has the form

$$U(x, \omega) \simeq \sum_{i=0}^{n_{PC}} \sum_{j=1}^{n_e} u_{i,j} \psi_i(G(\omega)) \zeta_j(x) = \sum_{j=1}^{n_e} \left[ \sum_{i=0}^{n_{PC}} u_{i,j} \psi_i(G(\omega)) \right] \zeta_j(x) \quad (9.90)$$

where  $(x, \omega) \in D \times \Omega$  and  $\{u_{i,j}\}$  are coefficients that need to be determined and  $\{\zeta_j(x)\}$  can be interpolators corresponding to a finite element partition of  $D$  with  $n_e$  nodes. The square brackets in this equation are truncated PC representations for the solutions at the nodes of the finite element partition of  $D$ .

The weak form of (9.53) in (9.54) with  $U(x, \omega)$  in (9.90) and  $\{\zeta_k \psi_l\}$  in place of  $W$  becomes

$$\sum_{j=1}^{n_e} \sum_{i=0}^{n_{PC}} u_{i,j} \mathcal{B}(\zeta_j \psi_i, \zeta_k \psi_l) = \langle f, \zeta_k \psi_l \rangle, \quad k = 1, \dots, n_e, l = 0, 1, \dots, n_{PC}, \quad (9.91)$$

where  $a$  in the expression of  $\mathcal{B}$  and  $f$  are replaced by their approximations in (9.89). The latter equalities constitute a system of equations for the unknown coefficients  $\{u_{i,j}\}$  in the approximate representation of  $U(x, \omega)$ . Once  $\{u_{i,j}\}$  have been obtained, the expression of  $U(x, \omega)$  in (9.90) becomes known, so that its probabilistic characteristics can be estimated efficiently by Monte Carlo simulation.

#### 9.4.8.3 Solution Existence and Properties

Let  $U(x, \omega)$  be the solution of the stochastic boundary value problem in (9.53) defined on the product space  $(D \times \Omega, \mathcal{B}(D) \times \mathcal{F}, \lambda \times P)$ , where  $D$  is a bounded

subset of  $\mathbb{R}^d$  with boundary  $\partial D$  (Sect. 9.4.3). It is assumed that (1) the random fields  $a(x)$  and  $f(x)$  are measurable in both arguments, (2) the correlation functions of  $a(x)$  and  $f(x)$  are continuous in  $D \times D$ , (3) the random field  $a(x)$  is uniformly bounded  $P$ -a.s. in  $\overline{D}$ , that is,  $P(a(x) \in [\alpha, \beta], x \in \overline{D}) = 1$  for  $0 < \alpha \leq \beta < \infty$ , (4) the field  $a(x)$  has uniformly bounded and continuous derivative, that is,  $P(a \in C^1(\overline{D}), \sup_{x \in \overline{D}} |\nabla a(x)| < c) = 1$  for a constant  $c > 0$ , and (5) the source term  $f$  is square integrable  $P$ -a.s. in  $D$ , that is,  $E[\int_D f(x)^2 dx] < \infty$  [20].

The first assumption allows the interchange of the order of integration in the physical and probability spaces by Fubini's theorem (Sect. 2.6). Assumption (2) guarantees the existence of KL expansions for the random fields  $a(x)$  and  $f(x)$  (Sect. 3.6.5). The bilinear form  $\mathcal{B}(U, W)$ ,  $V, W \in \mathcal{W}(D, \Omega)$ , in (9.56) is continuous and elliptic by assumption (3). Assumption (4) is needed to ensure the regularity of solution  $U(x)$ . The linear functional  $\langle f, W \rangle_{L^2(D \times \Omega)}$  of  $W$  is bounded by assumption (5). Under these assumptions, the weak form of (9.53) in (9.54) admits a unique solution by the Lax–Milgram theorem (Sect. B.4.2). Moreover, bounds can be established on the discrepancy between solutions of distinct elliptic stochastic boundary value problems satisfying the above assumptions (Theorem 9.4). Alternative bounds holding under various assumptions can be found in [28, 20].

We have seen that the discretization of the probability space involves the representation of the random fields  $a(x)$  and  $f(x)$  in the definition of (9.53) by parametric models depending on a random vector  $Z = (Z_1, \dots, Z_m)$ ,  $m < \infty$ , and of  $Z$  by truncated polynomial chaos series. It is common to assume that the coordinates  $\{Z_k\}$  of  $Z$  are independent random variables taking values in bounded intervals  $\Gamma_k = Z_k(\Omega)$ , so that the support of  $Z$  is the rectangle  $\Gamma = \times_{k=1}^m \Gamma_k$  in  $\mathbb{R}^m$ . Accordingly, the stochastic variational problem in (9.54) can be viewed as a deterministic variational problem

$$\int_{\Gamma} \langle a \nabla U, \nabla W \rangle_{L^2(D)} \rho(z) dz = \int_{\Gamma} \langle f, W \rangle_{L^2(D)} \rho(z) dz, \quad \forall W \in \mathcal{W}_{\rho}(D, \Omega), \quad (9.92)$$

where  $\rho(z)$  denotes the density of  $Z$ , the functions  $a, f$ , and  $U$  are viewed as depending on  $x$  and  $z$ , and  $\mathcal{W}_{\rho}(D, \Omega)$  is the space in (9.55) restricted to  $\mathbb{R}^m$ -valued random variables with range  $\Gamma$  that are mean square integrable with respect to  $\rho$ .

It is assumed without loss of generality that  $\Gamma_k$  are unit intervals on the real line and that the law of  $Z$  coincides with the Lebesgue measure, that is, the coordinates  $\{Z_k\}$  of  $Z$  are independent  $U(-1/2, 1/2)$  variables. Let  $\mathcal{P}_{r_k}$  denote the subspace of  $L^2(\Gamma_k)$  spanned by polynomials with degree at most  $r_k$  and let  $\mathcal{P}_r = \times_{k=1}^m \mathcal{P}_{r_k}$  be a polynomial subspace of  $L^2(\Gamma)$ . Denote by  $n_{\text{PC}}$  the dimension of  $\mathcal{P}_r$  and let  $U^{(m, n_{\text{PC}})}$  be the solution of (9.53) with a parametric model  $a(x, Z)$  in place of  $a(x)$  and  $Z$  approximated by a member of  $\mathcal{P}_r$ . Denote by  $U^{(m)}(x)$  the solution of (9.53) with  $a(x, Z)$  and  $f(x, Z)$  in place of  $a(x)$  and  $f(x)$ . The function  $U^{(m, n_{\text{PC}})}(x)$  is  $U^{(m)}(x)$  for  $a(x, Z)$  and  $f(x, Z)$  in which  $Z$  is approximated by polynomial chaoses up to degree  $n_{\text{PC}}$ . The rate of convergence  $U^{(m, n_{\text{PC}})}$  to  $U^{(m)}$  is given by the following statement ([21], Theorem 4.9).

**Theorem 9.7** *If  $\alpha \leq a(x) \leq \beta$ ,  $0 < \alpha \leq \beta < \infty$ ,  $P$ -a.s.,  $P$  coincides with the Lebesgue measure,  $\Gamma = [-1/2, 1/2]^m$ , and the correlation function of  $a(x)$  is piecewise analytic, then*

$$\|U^{(m)} - U^{(m, n_{\text{PC}})}\|_{H_0^1(D) \times L^2(\Gamma)} \leq \exp(-c_1 c_2^{-1/d} (\log(n_{\text{PC}}))^{1/d}) \quad (9.93)$$

where  $c_1, c_2 > 0$  are constants such that  $n_{\text{PC}} \leq \exp(\beta m)$ .

The bound in (9.93) shows that, for a given parametric model  $a(x, Z)$  of  $a(x)$  depending on  $m$  random variables, the accuracy of  $U^{(m, n_{\text{PC}})}(x)$  increases with the dimension  $n_{\text{PC}}$  of the polynomial chaos representation and decreases with the dimension  $d$  of the physical space. We conclude this section with a numerical example illustrating the implementation of the stochastic Galerkin method and its potential limitations.

*Example 9.17* Let  $U(x)$  be the solution of the stochastic partial differential equation (9.53) in Example 9.8 with  $a(x)$  in (9.62), (9.63), and (9.64). We apply the stochastic Galerkin method to characterize  $U(x)$ . The implementation of this method requires to discretize the probability and physical spaces. The discretization of the probability space is commonly accomplished by replacing the random field  $a(x)$  in (9.53) with the linear parametric model  $a^{(n, \zeta)}(x)$  in (9.66), that can be given in the form

$$a^{(n, \zeta)}(x, \omega) = a_0 + \sum_{r=1}^m \alpha_r(x) Z_r(\omega), \quad (9.94)$$

where  $m = 2n$ ,  $\{Z_r\}$  and  $\alpha_r(x)$  correspond to the random variables and the deterministic functions in the expression of  $a^{(n, \zeta)}(x)$ . For example,  $Z_1 = C_1$ ,  $Z_2 = D_1$ ,  $\alpha_1(x) = \zeta \sigma_k \cos(v^{(k)} \cdot x)$ , and  $\alpha_2(x) = \zeta \sigma_k \sin(v^{(k)} \cdot x)$ . The random variables  $\{C_k, D_k\}$  are independent  $U(-\sqrt{3}, \sqrt{3})$ , so that they have mean 0 and variance 1, and  $\zeta = 0.3286$  is a scale factor such that most samples of  $a^{(n, \zeta)}(x)$  are positive (Example 9.8). The random variables  $\{Z_r\}$  in (9.94) are approximate by

$$Z_r(\omega) \simeq \sum_{k=1}^{n_{\text{PC}}} a_{r,k} \tilde{H}_k(G_r(\omega)), \quad (9.95)$$

where  $n_{\text{PC}} \geq 0$  is an integer denoting the largest degree of polynomial chaoses considered for solution,  $\{H_k(x)\}$  denote Hermite polynomials (Sect. B.6.1),  $\tilde{H}_k(x) = H_k(x)/\sqrt{k!}$  so that  $E[\tilde{H}_k(G_r) \tilde{H}_l(G_r)] = \delta_{kl}$ ,  $G_r \sim N(0, 1)$ , and the coefficients in the expression of  $Z_r(\omega)$  are  $a_{r,k} = E[Z_r \tilde{H}_k(G_k)]$ . Note that  $a^{(n, \zeta)}(x)$  in (9.94) with  $Z_r$  in (9.95) takes the form in (9.89).

The discretization of the physical space has been discussed in Sect. 9.4.8.2. It represents the solution  $U(x, \omega)$  as a member of a finite dimensional space spanned by a family of specified functions  $(\zeta_1(x), \dots, \zeta_{n_e}(x))$  for almost all  $\omega \in \Omega$ , that is,

$$U(x, \omega) \simeq \sum_{i=1}^{n_e} U_i(\omega) \zeta_i(x), \quad x \in D, \quad \omega \in \Omega. \quad (9.96)$$

The coefficients  $\{U_i\}$  are functions of the random vector  $Z = (Z_1, \dots, Z_m)$ , which is mapped by (9.95) into a Gaussian vector  $G = (G_1, \dots, G_m)$  with independent  $N(0, 1)$  coordinates. The random variables  $\{U_i\}$  are represented approximately by their projections on the subspace of  $\mathcal{W}(D, \Omega)$  spanned by the polynomial chaoses

$$\left\{ \prod_{r=1}^m \tilde{H}_{k_r}(G_r), k_1 + \dots + k_m = 0, 1, \dots, n_{\text{PC}}, k_1, \dots, k_m \geq 0 \right\}, \quad (9.97)$$

up to degree  $n_{\text{PC}}$  (Sect. B.6.2), that is,

$$U_i(\omega) \simeq \sum_{k_1 + \dots + k_m = 0, k_1, \dots, k_m \geq 0}^{n_{\text{PC}}} u_{i, k_1, \dots, k_m} \prod_{r=1}^m \tilde{H}_{k_r}(G_r(\omega)), \quad (9.98)$$

where  $\{u_{i, k_1, \dots, k_m}\}$  are real-valued coefficients. If these coefficients are known, properties of  $U(x)$  can be calculated simply from (9.96) and (9.98) by Monte Carlo simulation.

We replace the solution of (9.53) with that of

$$\sum_{l=1}^2 \frac{\partial a(x)}{\partial x_l} \frac{\partial V(x)}{\partial x_l} + a(x) \Delta V(x) = -\frac{1}{l_1} \frac{\partial a(x)}{\partial x_1}, \quad (9.99)$$

where  $V(x)$  and  $U(x)$  are related by  $V(x) = U(x) - x_1/l_1$ . This equation is preferred since it has homogeneous Dirichlet boundary conditions. The boundary conditions for (9.99) are  $V(0, x_2) = 0$  and  $V(l_1, x_2) = 0$  for  $x_2 \in (0, l_2)$  and  $\partial V(x_1, 0)/\partial x_2 = \partial V(x_1, l_2)/\partial x_2 = 0$  for  $x_1 \in (0, l_1)$ . The solution  $V(x, \omega)$  of (9.99) admits the same representation as  $U(x, \omega)$ , that is,

$$V(x, \omega) \simeq \sum_{i=1}^{n_e} V_i(\omega) \zeta_i(x), \quad x \in D, \quad \omega \in \Omega \text{ and} \\ V_i(\omega) \simeq \sum_{k_1 + \dots + k_m = 0, k_1, \dots, k_m \geq 0}^{n_{\text{PC}}} v_{i, k_1, \dots, k_m} \prod_{r=1}^m \tilde{H}_{k_r}(G_r(\omega)), \quad (9.100)$$

where  $\{v_{i, k_1, \dots, k_m}\}$  are real-valued coefficients that need to be determined.

The weak form of (9.99) results by using the representations in (9.94), (9.95), and (9.100) for  $a^{(n, \zeta)}(x, \omega)$  and  $V(x, \omega)$ , multiplying the resulting form of (9.99) by  $\zeta_j(x) \prod_{q=1}^m \tilde{H}_{s_q}(G_q)$  for  $j = 1, \dots, n_e$  and  $s_1 + \dots + s_m = 0, 1, \dots, n_{\text{PC}}$  such that  $s_1, \dots, s_m \geq 0$ , and integrating over  $D \times \Omega$ . We obtain

$$\sum_{i=1}^{n_e} \sum_{k_1, \dots, k_m} v_{i, k_1, \dots, k_m} \left\{ \sum_{r=1}^m \sum_{l=1}^2 \sum_{k=0}^{n_{\text{PC}}} a_{r, k} \int_D \frac{\partial \alpha_r(x)}{\partial x_l} \frac{\partial \zeta_i(x)}{\partial x_l} \zeta_j(x) dx \right. \\ \times E \left[ \tilde{H}_k(G_r) \prod_{p=1}^m \tilde{H}_{k_p}(G_p) \prod_{q=1}^m \tilde{H}_{s_q}(G_q) \right] \\ \left. + a_0 \int_D \Delta \zeta_i(x) \zeta_j(x) dx E \left[ \prod_{p=1}^m \tilde{H}_{k_p}(G_p) \prod_{q=1}^m \tilde{H}_{s_q}(G_q) \right] \right\}$$



$$\begin{aligned}
& + \sum_{r=1}^m \sum_{l=1}^2 \sum_{k=0}^{n_{\text{PC}}} a_{r,k} \int_D \alpha_r(x) \Delta \zeta_j(x) \zeta_j(x) dx \\
& \times E \left[ \tilde{H}_k(G_r) \prod_{p=1}^m \tilde{H}_{k_p}(G_p) \prod_{q=1}^m \tilde{H}_{s_q}(G_q) \right] \Big\} \\
& = -\frac{1}{l_1} \sum_{r=1}^m \sum_{k=0}^{n_{\text{PC}}} a_{r,k} \int_D \frac{\partial \alpha_r(x)}{\partial x_1} \zeta_j(x) dx E \left[ \tilde{H}_k(G_r) \prod_{q=1}^m \tilde{H}_{s_q}(G_q) \right], \quad (9.101)
\end{aligned}$$

where the symbol  $\Sigma$  means  $\sum_{k_1+\dots+k_m=0, k_1, \dots, k_m \geq 0}^{n_{\text{PC}}}$ . This equation written for  $j = 1, \dots, n_e$  and  $s_1 + \dots + s_m = 0, 1, \dots, n_{\text{PC}}$  such that  $s_1, \dots, s_m \geq 0$  supplies a system of algebraic equations for  $v_{i;k_1, \dots, k_m}$ . The solution of this system of algebraic equation and (9.100) define an approximate representation for  $V(x, \omega)$  that can be used to calculate properties of  $V(x)$  and  $U(x)$ .

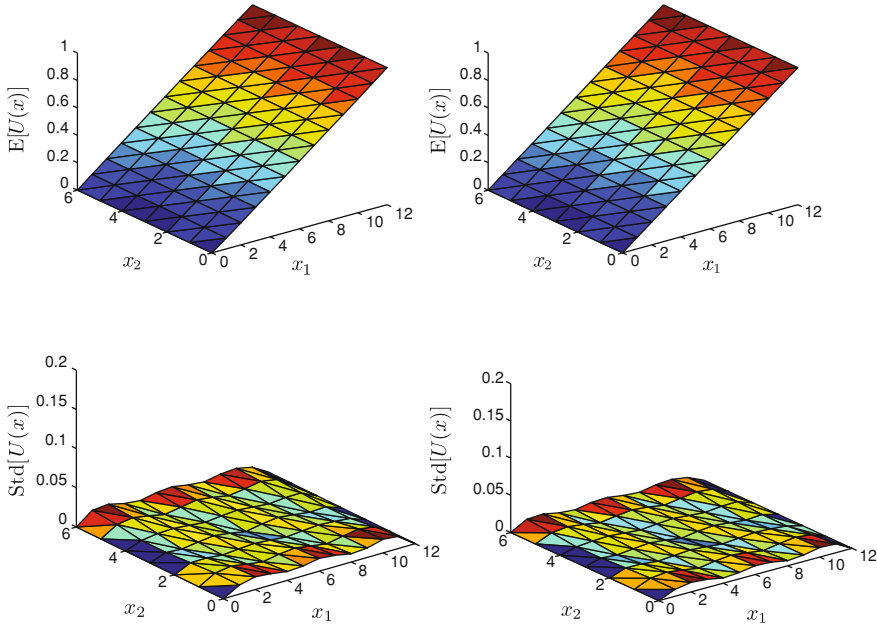
Numerical results have been obtained for  $n_{\text{PC}} = 2$ ,  $n = 4$ ,  $m = 8$ , and  $\zeta = 0.3286$ . The coordinates of the frequencies  $v^{(k)}$ ,  $k = 1, \dots, 4$ , in the representation of  $a^{(n, \zeta)}$  given by (9.94) are  $v_1^{(k)} = \pm 1.5$  and  $v_2^{(k)} = \pm 1.5$ , and  $\sigma_k^2$  are the volumes under the spectral density in Fig. 9.5 for  $\{(v_1, v_2) : v_1 > 0, v_2 > 0\}$ ,  $\{(v_1, v_2) : v_1 > 0, v_2 < 0\}$ ,  $\{(v_1, v_2) : v_1 < 0, v_2 > 0\}$ , and  $\{(v_1, v_2) : v_1 < 0, v_2 < 0\}$ . This representation is one of the cases considered in Example 9.8, and illustrated in Figs. 9.8 and 9.9 (top right panel).

We have used polynomial chaoses of low degree and just a few harmonics to minimize calculations. For our selection  $n_{\text{PC}} = 2$  and  $n = 4$ , a system with 4095 equations has been obtained and solved to find the coefficients  $v_{i;k_1, \dots, k_m}$  in the expansion of  $V(x, \omega)$ . Larger values for  $n_{\text{PC}}$  and  $n$  may not be justified since the conditions that  $a^{(n, \zeta)}$  must satisfy yield solutions  $U(x)$  with very small variance.

The left and right top panels in Fig. 9.29 show expectations of  $U(x)$  obtained by Monte Carlo simulation and the stochastic Galerkin method. The bottom left and right panels show standard deviations of  $U(x)$  by Monte Carlo and stochastic Galerkin. The Monte Carlo estimates of the mean and standard deviation of  $U(x)$  are based on 800 samples. The mean and standard deviation of  $U(x)$  delivered by the stochastic Galerkin method are satisfactory.  $\diamond$

### 9.4.9 Stochastic Collocation Method

Consider the stochastic initial-boundary value problem in (9.33). Suppose the random fields in this equation are replaced by parametric models depending on a random vector  $Z$  defined on a probability space  $(\Omega, \mathcal{F}, P)$ . The representations  $a(x, Z)$  in (9.84) or  $a_T^{(n)}(x)$  in (9.87) for the random conductivity  $a(x)$  in (9.62) are examples of parametric models. The solution  $U(x, t, Z)$  of this version of (9.33) is a deterministic function of the spatial and temporal coordinates  $(x, t)$  and the random vector  $Z$ , so that  $U(x, t, Z)$  is a parametric model.



**Fig. 9.29** Expectations of  $U(x)$  by Monte Carlo (*top left panel*) and stochastic Galerkin (*top right panel*) and standard deviations of  $U(x)$  by Monte Carlo (*bottom left panel*) and stochastic Galerkin (*bottom right panel*)

The construction of collocation solutions involves three steps. First, collocation points  $z^{(i)}$ ,  $i = 1, \dots, n_c$ , need to be selected in the image  $\Gamma = Z(\Omega)$  of  $Z$ . Second,  $n_c$  deterministic analyses need to be performed to find the solutions  $u^{(i)}(x, t) = U(x, t, z^{(i)})$  of (9.33) with  $Z = z^{(i)}$ . Finite element or any other method can be used to calculate these solutions. Third, approximations  $\tilde{U}(x, t, Z)$  need to be developed for  $U(x, t, Z)$  from the deterministic solutions  $u^{(i)}(x, t)$ ,  $i = 1, \dots, n_c$ . Interpolation polynomials are commonly used to construct  $\tilde{U}(x, t, Z)$ . Probabilistic characteristic of  $\tilde{U}(x, t, Z)$  can be obtained efficiently by Monte Carlo simulation, quadratures, or other methods since the functional form of  $\tilde{U}(x, t, Z)$  is known.

The approximate solution  $\tilde{U}(x, t, Z)$  can be viewed as a response surface for  $U(x, t, Z)$  developed over the range  $\Gamma = Z(\Omega)$  of  $Z$ . The performance of the collocation solution  $\tilde{U}(x, t, Z)$  depends on the accuracy of the finite element algorithm used to calculate  $\{u^{(i)}(x, t)\}$ , the number and the location of collocation points, the type of interpolation polynomials, and the properties of  $U(x, t, Z)$ , which are not known. The accuracy of the algorithm for calculating the deterministic solutions  $\{u^{(i)}(x, t)\}$  is the only item on which we have full control. The number of collocation points is primarily dictated by computation time, that is, the time needed to compute a deterministic solution  $u^{(i)}(x, t)$ . The location of collocation points is usually selected by algorithms that are unrelated to the distribution of  $Z$ . The dimension of  $Z$  must

be kept relatively small if full tensor grid is used to sample  $Z$ , as illustrated by the following example.

*Example 9.18* Let  $U(x)$  be the solution of  $d(a(x)dU(x)/dx)/dx = 0$ ,  $x \in (0, l)$ ,  $0 < l < \infty$ , with boundary conditions  $U(0) = 0$  and  $U(l) = 1$ , where  $a(x)$  is a strictly positive and bounded random field with continuous samples. We have

$$U(x) = \int_0^x K(u) du / \int_0^l K(v) dv, \quad x \in D = (0, l), \quad (9.102)$$

with the notation  $K(x) = 1/a(x)$ . Suppose  $K(x)$  is the Beta translation random field

$$K(x) = \alpha + (\beta - \alpha) F_{\text{Beta}(p,q)}^{-1} \circ \Phi(G(x)), \quad x \in D = (0, l), \quad (9.103)$$

where  $F_{\text{Beta}(p,q)}$  denotes the distribution of a standard Beta variable with shape parameters  $(p, q)$ ,  $0 < \alpha < \beta < \infty$  are constants, and  $G(x)$  is a homogeneous Gaussian field with mean 0, variance 1, and one-sided spectral density  $g(v)$ ,  $v \geq 0$ .

The Monte Carlo solution for this problem requires to generate samples of  $G(x)$ , calculate corresponding samples of  $K(x)$  and  $U(x)$  from (9.103) and (9.104), and estimate properties of  $U(x)$  from its samples. Samples of  $G(x)$  can be generated from, for example, the discrete spectral representation (Sect. 3.8.1.2)

$$G^{(n)}(x) = \sum_{k=1}^n \sigma_k^{(n)} (A_k \cos(v_k^{(n)} x) + B_k \sin(v_k^{(n)} x)), \quad x \in (0, l), \quad (9.104)$$

where  $n \geq 1$  is an integer,  $0 < \bar{v} < \infty$  denotes a cutoff frequency,  $v_k^{(n)} = (k - 1/2)\bar{v}/n$ ,  $(\sigma_k^{(n)})^2$  is the area under  $\alpha g(v)$  in the interval  $(v_k^{(n)} - \bar{v}/(2n), v_k^{(n)} + \bar{v}/(2n))$ ,  $k = 1, \dots, n$ , and  $\alpha = 1/\sum_{k=1}^n (\sigma_k^{(n)})^2$  is a scaling factor such that  $G^{(n)}(x)$  has unit variance.

The stochastic collocation method can use a parametric model for  $K(x)$  given by (9.103) with  $G(x)$  replaced by

$$G^{(n)}(x) = \sum_{k=1}^n \sigma_k^{(n)} (\Phi^{-1}(U_k) \cos(v_k^{(n)} x) + \Phi^{-1}(V_k) \sin(v_k^{(n)} x)), \quad x \in (0, l), \quad (9.105)$$

with the notations in (9.104), where  $\{U_k, V_k\}$  are independent  $U(0, 1)$  variables. The resulting representation  $K^{(n)}(x)$  for  $K(x)$  is a nonlinear parametric model depending on the  $m$ -dimensional vector  $Z = (U_1, V_1, \dots, U_n, V_n)$ ,  $m = 2n$ , so that the collocation solution is a function  $U(x, Z)$  of  $x \in D$  and  $Z$ . The implementation of the collocation method requires to select collocation points  $z^{(i)}$ ,  $i = 1, \dots, n_c$ , in the image  $\Gamma = Z(\Omega) = [0, 1]^m$  of  $Z$ , calculate deterministic solutions  $u^{(i)}(x)$  corresponding to  $Z = z^{(i)}$ , construct approximations  $\tilde{U}(x, Z)$  for  $U(x)$  by interpolating

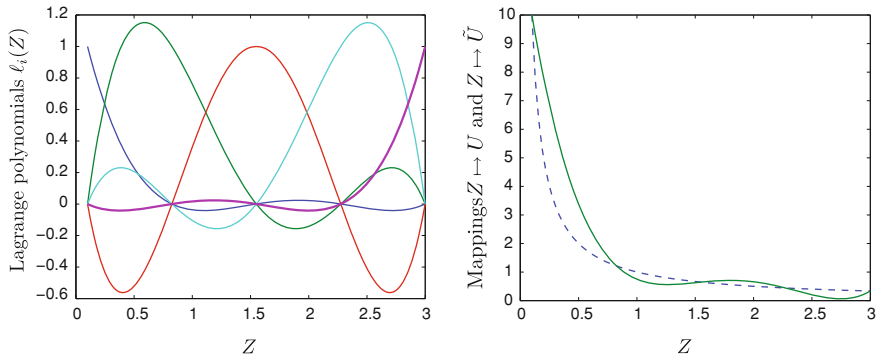
over  $\Gamma$  between  $\{u^{(i)}(x)\}$ . There is no general methodology for selecting collocation points in an optimal manner. If factorial design with two levels for each coordinate of  $Z$  is used, there will be  $n_c = 2^m$  collocation points, for example, 16, 1024, and 1048576 points for  $m = 4, 10$ , and 20, respectively. This suggests that collocation solutions based on a full tensor grid are not feasible in applications since  $Z$  is likely to have a relatively large dimension.  $\diamond$

An additional difficulty relates to the accuracy of polynomial interpolations that deteriorates with the dimension of  $Z$ . We summarize essential properties of polynomial approximations for functions defined on intervals of the real line and extend these results to functions defined on bounded subsets of  $\mathbb{R}^d$ ,  $d > 1$ .

Interpolation polynomials commonly used to construct collocation solutions are defined in [Sect. B.6](#) within the framework of the Sturm-Liouville differential equations. Following facts on the accuracy of polynomials approximations for real-valued functions are summarized for convenience. Let  $I$  be a bounded interval of the real line,  $w : I \rightarrow \mathbb{R}$  a positive weighting function, and  $L_w^2(I) = \{f : I \rightarrow \mathbb{R} : \int_I f(x)^2 w(x) dx < \infty\}$ . Note that  $\langle f, g \rangle_{L_w^2(I)} = \int_I f(x)g(x)w(x) dx$ ,  $f, g \in L_w^2(I)$ , is an inner product on  $L_w^2(I)$  inducing the norm  $\|f\|_{L_w^2(I)}^2 = \int_I f(x)^2 w(x) dx$  on this space. Let  $\{\varphi_k(x), k = 0, 1, \dots, n\}$  be orthonormal polynomials  $\mathcal{P}_n(I)$  with degree at most  $n$ , so that  $\langle \varphi_k, \varphi_l \rangle_{L_w^2(I)} = \delta_{kl}$  and the projection of  $f \in L_w^2(I)$  on the space spanned by these polynomials is  $\pi_n(f) = \sum_{k=0}^n \langle f, \varphi_k \rangle_{L_w^2(I)} \varphi_k(x)$ . Then  $\|f - \pi_n(f)\|_{L_w^2(I)} \leq \inf_{p \in \mathcal{P}_n(I)} \|f - p\|_{L_w^2(I)}$  ([\[36\]](#), Theorem 3.3, and [Sect. B.4.1](#) in this book). It can also be shown that  $\|f - \pi_n(f)\|_{L_w^2(I)} \rightarrow 0$  as  $n \rightarrow \infty$  ([\[36\]](#), Theorem 3.5). The rate of convergence of  $\pi_n(f)$  to  $f$  depends on the smoothness of  $f$ . For example, if  $f$  and its first  $p \geq 0$  derivatives are in  $L_w^2(I)$  and  $w(x) = 1$ , that is,  $f$  is a member of the Sobolev space  $H^p(I)$  defined by [\(9.36\)](#),  $I = [-1, 1]$ , and  $\varphi_k(x)$  are Legendre polynomials, then  $\|f - \pi_n(f)\|_{L_w^2(I)} \leq cn^{-p} \|f\|_{H^p(I)}$ , where  $c > 0$  is a constant and  $\|\cdot\|_{H^p(I)}$  denotes the norm in [\(9.37\)](#) ([\[36\]](#), Theorem 3.5). Hence, the error of a polynomial approximation  $\pi_n(f)$  is of order  $O(n^{-p})$ , so that it relates directly to the degree of smoothness of  $f$ .

Consider now the solution  $U(x, t, Z)$  of [\(9.33\)](#) and suppose that  $n_{c,1}$  collocation points are used along each coordinate of  $Z$ , so that the full tensor grid has  $n_c = n_{c,1}^m$  collocation points, where  $m = 2n$  denotes the dimension of  $Z$ . The error of a polynomial approximation  $\pi_{n_{c,1}}(Z_r)$  for  $U(x, t, Z)$  viewed as a function of  $Z_r$  and all other arguments arbitrary but fixed is of order  $O(n_{c,1}^{-p})$ , where  $p$  denotes the order of differentiation of  $U(x, t, Z)$  with respect to  $Z_r$ . Hence, the error of the polynomial approximation  $p(Z) = \prod_{r=1}^m \pi_{n_{c,1}}(Z_r)$  for  $U(x, t, Z)$  constructed on the full tensor grid is of order  $O(n_{c,1}^{-p/m})$ , so that it increases rapidly with the dimension of  $Z$ . The accuracy of collocation solutions depends also on the location of collocation points, as illustrated by the following example.

*Example 9.19* Let  $U$  be the solution of the stochastic algebraic equation  $ZU = 1$ , where  $Z$  is a Beta random variable with range  $(\alpha, \beta)$ ,  $0 < \alpha \leq \beta < \infty$ , and shape parameters  $(p, q)$ . The collocation solution has the expression



**Fig. 9.30** Lagrange polynomials  $\ell_i(Z)$  (left panel) and exact and approximate mappings (right panel)

$$\tilde{U}(Z) = \sum_{i=0}^n \frac{1}{\alpha + i(\beta - \alpha)/n} \ell_i(Z). \quad (9.106)$$

where  $\ell_i(Z) = \prod_{j=0, j \neq i}^n (Z - z^{(j)}) / (z^{(i)} - z^{(j)})$  are Lagrange interpolation polynomials (Example 8.16). The left panel in Fig. 9.30 shows Lagrange interpolation polynomials  $\ell_i(Z)$  constructed on  $n_c = 5$  equally spaced collocation points, two of which are  $\alpha$  and  $\beta$ . The solid and the dotted lines in the right panel of the figure are the exact mapping  $Z \mapsto U(Z) = 1/Z$  and the mapping  $Z \mapsto \tilde{U}(Z)$  in (9.106). The approximate mapping provides a reasonable overall representation for  $U(Z)$ . The table,

$(p = 1/2, q = 3)$	30.96;	55.37;	69.12;	77.07;	82.06;	85.45
$(p = 1, q = 1)$	25.91;	100.12;	161.37;	200.43;	226.68;	245.72
$(p = 2, q = 5)$	35.18;	135.23;	287.36;	452.44;	605.25;	738.50,

gives errors  $100(E[\tilde{U}(Z)^r] - E[U(Z)^r])/E[U(Z)^r]$  in percentages for the first six moments of  $\tilde{U}(Z)$  relative to corresponding exact moments for  $[\alpha, \beta] = [0.1, 3]$  and several values of the shape parameters  $(p, q)$ . The errors depend strongly on both properties of  $Z$  and the quality of the approximate mapping  $Z \mapsto \tilde{U}(Z)$  in the intervals of likely values of  $Z$ . For example, errors are significant for  $(p = 2, q = 5)$  since the density of  $Z$  is skewed to the left for these parameters and the mapping  $Z \mapsto \tilde{U}(Z)$  is the least accurate in the interval  $(0.2, 0.7)$ . The errors for  $(p = 2, q = 5)$  are reduced to  $-7.74, 23.01, 70.88, 129.00, 184.12,$  and  $233.12$  if the collocation points  $(0.1, 0.5, 1, 2, 3)$  are used, suggesting that the collocation grid needs to be refined in the subsets of  $\Gamma = Z(\Omega)$  in which  $Z$  resides with relatively high probabilities and  $\tilde{U}(Z)$  is less accurate. Unfortunately, this scheme cannot be implemented fully since it requires information on the behavior of  $\tilde{U}(Z)$  as a function of  $Z$ , and this information is not available.  $\diamond$

We have seen that the accuracy of collocation solutions depends on the quality of the parametric models used to represent the random fields in the definition of stochastic equations, for example, the parametric model  $a^{(n)}(x)$  of  $a(x)$  in (9.65), the distribution postulated for the  $m$ -dimensional random vector  $Z$  in the definition of parametric models, the accuracy of deterministic solvers, and the performance of polynomial interpolation. Let  $U^{(n)}(x)$  be a collocation solution corresponding to a parametric model  $a^{(n)}(x)$  of  $a(x)$ ,  $\pi_h U^{(n)}$  projections of  $U^{(n)}(x)$  on, for example, the finite element space, and  $\mathcal{P}(\pi_h U^{(n)})$  polynomials interpolations constructed on the finite element solutions  $\pi_h U^{(n)}$ . We have

$$\|U - \mathcal{P}(\pi_h U^{(n)})\| \leq \|U - U^{(n)}\| + \|U^{(n)} - \pi_h U^{(n)}\| + \|\pi_h U^{(n)} - \mathcal{P}(\pi_h U^{(n)})\| \quad (9.107)$$

in some norm. Bounds on the first two norms result from Theorem 9.4 and (9.59). A relatively simple bound on the third norm can be obtained under the assumptions that (1) the solution  $\pi_h U^{(n)}(x, y)$  viewed as a function of  $(x, y) \in D \times \Gamma$ ,  $\Gamma = Z(\Omega)$  admits an analytic extension in  $\Sigma_j = \{z \in \mathbb{C} : \text{dist}(z, \Gamma_j) \leq \alpha_j\}$ ,  $j = 1, \dots, m$ ,  $\alpha_j > 0$ , for each coordinate of  $y = (y_1, \dots, y_j, \dots, y_m)$ , (2) the density of  $Z$  is uniform, and (3) the range  $\Gamma$  of  $Z$  is partitioned in finite elements with mesh size  $h$ . Under these assumptions, the discrepancy between  $\pi_h U^{(n)}$  and its approximation  $\mathcal{P}(\pi_h U^{(n)})$  given by Lagrange polynomials satisfies the inequality

$$\|\pi_h U^{(n)} - \mathcal{P}(\pi_h U^{(n)})\| \leq c \sum_{j=1}^d \exp(-r_j(h)p_j), \quad (9.108)$$

where  $r_j(h) = \log((\alpha_j/(2h))[1 + \sqrt{1 + h^2/\alpha_j^2}])$  and  $c > 0$  denotes a constant [29, 37]. The accuracy of polynomial approximation increases as the mesh is refined, that is, the mesh size  $h$  is reduced. Note that finer meshes are needed for larger vectors  $Z$ . Alternative bounds on the accuracy of polynomial interpolation in the probability space can be found in [38] and [36] (Chap. 7) for both full tensor and sparse collocation grids.

It turns out that the full tensor grid can be reduced significantly while preserving the accuracy of resulting interpolation formulas by using sparse grid collocation techniques based on the Smolyak algorithm. For the construction, properties, and implementation of the Smolyak algorithm the reader is directed to [29, 37, 39, 38]. Additional information on sparse grid collocation can be found in [36] (Sect. 7.2.2). In contrast to collocation solutions based on full tensor grids that are feasible for problems depending on a small number of random parameters [31], sparse grid collocation solutions can be used to solve relatively large stochastic problems [39].

We conclude this section with the solution of the stochastic partial differential equation in Example 9.17 by the stochastic collocation method. The analysis uses the same linear parametric model for  $a(x)$  as in this example.

*Example 9.20* Let  $U(x)$  be the solution of the stochastic partial differential equation (9.53) in Example 9.8, where  $a(x)$  is approximated by the linear parametric model

$a^{(n,\zeta)}(x)$  in (9.66). Let  $Z = (C_1, D_1, \dots, C_n, D_n)$  be the  $\mathbb{R}^m$ -valued random variable collecting the random coefficients in the expression of  $a^{(n,\zeta)}(x)$ , where  $m = 2n$ . Denote by  $\Gamma_r = Z_r(\Omega)$ ,  $r = 1, \dots, m$ , the images of the coordinates of  $Z$ , so that  $\Gamma = Z(\Omega) = \times_{r=1}^m \Gamma_r$ . Let  $(\eta_{r,1}, \dots, \eta_{r,n_{c,r}})$  be collocation points along coordinate  $r$  of  $Z$ , where  $n_{c,r} \geq 1$  is an integer. The collocation solution has the form

$$\tilde{U}(x) = \sum_{i=1}^{n_c} u^{(i)}(x) \ell^{(i)}(Z), \quad (9.109)$$

where  $n_c = \prod_{r=1}^m n_{c,r}$  denotes the number of collocation points,  $\{u^{(i)}(x)\}$  are solutions of (9.53) for  $Z$  set equal to the selected collocation points  $\eta^{(i)} \in \Gamma$ ,  $i = 1, \dots, n_c$ ,

$$\begin{aligned} \ell^{(i)}(Z) &= \prod_{r=1}^{2n} \ell_r(Z_r), \text{ and} \\ \ell_r(Z_r) &= \prod_{j=1, j \neq r}^{n_{c,r}} \frac{Z_r - \eta_{r,j}}{\eta_{r,r} - \eta_{r,j}} \end{aligned} \quad (9.110)$$

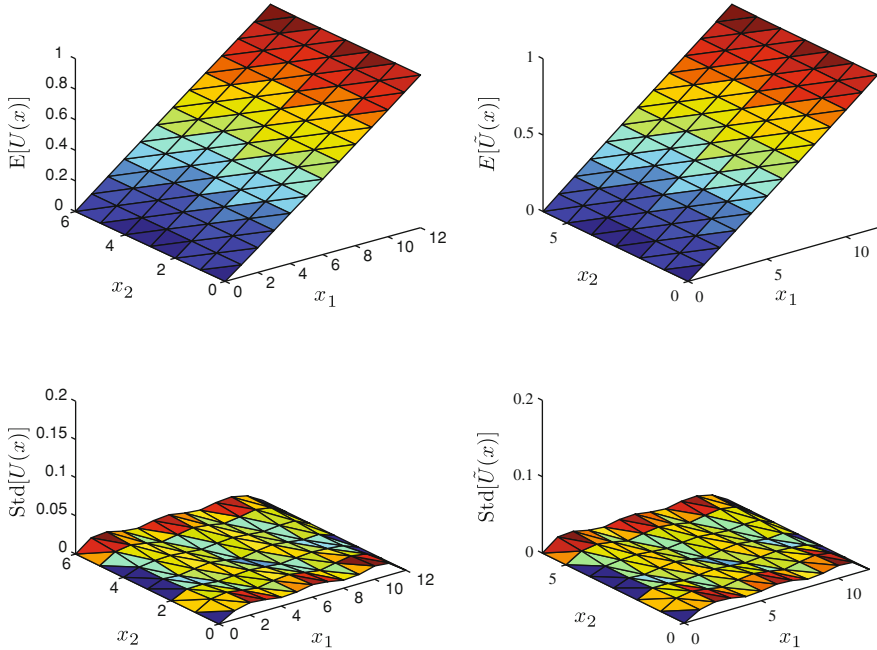
are Lagrange interpolation polynomials along the coordinates of  $Z$ .

Numerical results are for the same models and parameters as in Example 9.17 and collocation points  $\eta_{c,r} = \{-3/\sqrt{5}, 0, 3/\sqrt{5}\}$  along each coordinate of  $Z$ . Since  $Z$  has  $m = 2n = 8$  coordinates, the approximation of solution  $U(x)$  is based on  $3^m = 6561$  collocation points, so that it involves 6561 deterministic analyses. The left and right top panels in Fig. 9.31 show expectations of  $U(x)$  obtained by Monte Carlo simulation and the stochastic collocation method. The bottom left and right panels show standard deviations of  $U(x)$  by the Monte Carlo and stochastic collocation methods. The Monte Carlo estimates of the mean and standard deviation of  $U(x)$  are based on 800 samples. The mean and standard deviation of  $U(x)$  delivered by the stochastic collocation method are satisfactory.  $\diamond$

The stochastic Galerkin and collocation methods provide accurate approximations for the mean and standard deviation of the solution  $U(x)$  in Example 9.17. At a first glance, the remarkable performance of these methods may be surprising since both methods have used coarse representations of uncertainty. The low uncertainty in  $U(x)$  is the likely reason for the satisfactory performance of these methods.

## 9.5 Applied SPDEs: Small Uncertainty

Consider the stochastic elliptic boundary value problem in (9.53) with weak form given by (9.54). For simplicity, suppose the source term  $f(x)$  is deterministic and the coefficient  $a(x)$  can be approximated by the linear parametric model  $a(x, Z) =$



**Fig. 9.31** Expectations of  $U(x)$  by Monte Carlo (*top left panel*) and stochastic collocation (*top right panel*) and standard deviations of  $U(x)$  by Monte Carlo (*bottom left panel*) and stochastic collocation (*bottom right panel*)

$\sum_{k=1}^m a_k(x)Z_k$  in (9.84). Note that the solution  $U(x, Z)$  of (9.53) is also a parametric model. We assume that the stochastic boundary value problems under consideration admit unique solutions  $P$ -a.s.

### 9.5.1 Taylor Series

The first order Taylor approximation of  $U(x, Z)$  has the form

$$U(x, Z) \simeq U(x, \mu) + \sum_{k=1}^m V_k(x, \mu)(Z_k - \mu_k) = U(x, \mu) + \nabla U(x, \mu) \cdot (Z - \mu), \quad (9.111)$$

where  $\mu_k = E[Z_k]$ ,  $\mu = E[Z]$ ,  $V_k(x, Z) = \partial U(x, Z)/\partial Z_k$ , and  $m$  is the dimension of  $Z$ . To find properties of the above approximation of  $U(x, Z)$ , we need to calculate  $U(x, \mu)$  and  $\{V_k(x, \mu)\}$ . The solution of the deterministic version of (9.53) with  $Z = \mu$  gives  $U(x, \mu)$ . The gradient of  $U(x, Z)$  at  $Z = \mu$  can be obtained by differentiating (9.53) with respect to  $Z_j$ ,  $j = 1, \dots, m$ , which gives  $-\nabla \cdot (a_j(x)\nabla U(x, Z)) - \sum_{k=1}^m Z_k \nabla \cdot (a_k(x)\nabla V_j(x, Z)) = 0$ , and solving for



$\nabla U(x, Z)$  at  $Z = \mu$ . These operations give

$$-\sum_{k=1}^m \mu_k \nabla \cdot (a_k(x) \nabla V_j(x, \mu)) = \nabla \cdot (a_j(x) \nabla U(x, \mu)), \quad j = 1, \dots, m.$$

Since  $U(x, Z) = 0$  on  $\partial D$   $P$ -a.s., we can solve for  $V_j(x, \mu)$ .

The representation in (9.111) can be used to find properties of  $U(x, Z)$  approximately. For example, its mean and the correlation functions are

$$\begin{aligned} E[U(x, Z)] &\simeq U(x, \mu) \\ E[U(x, Z)U(y, Z)] &\simeq U(x, \mu)U(y, \mu) + \sum_{k,l=1}^m V_k(x, \mu)V_l(y, \mu)\gamma_{kl} \end{aligned} \quad (9.112)$$

where  $\{\gamma_{kl} = E[(Z_k - \mu_k)(Z_l - \mu_l)]\}$  is the covariance matrix of  $Z$  and  $x, y \in D$ .

*Example 9.21* The first order Taylor approximation of  $U(x, Z)$  defined by the stochastic boundary value problem  $-ZU''(x, Z) = 1$ ,  $x \in (0, 1)$  with the homogeneous boundary conditions  $U(0, Z) = U(1, Z) = 0$  is

$$U(x, Z) \simeq (x - x^2)/(2\mu) + (-x + x^2)(Z - \mu)/(2\mu^2), \quad (9.113)$$

since  $U(x, Z) = (x - x^2)/(2Z)$  and  $V(x, Z) = \partial U(x, Z)/\partial Z = (-x + x^2)/(2Z^2)$ . The mean and covariance functions for this approximation of  $U(x, Z)$  are

$$\begin{aligned} E[U(x, Z)] &\simeq \frac{x - x^2}{2\mu} \\ \text{Cov}[U(x, Z), U(y, Z)] &\simeq \frac{\text{Var}[Z]}{4\mu^4}(-x + x^2)(-y + y^2), \end{aligned} \quad (9.114)$$

and depend on only the first two moments of  $Z$ .  $\diamond$

### 9.5.2 Perturbation Series

Suppose the random variables in the parametric model  $a(x, Z) = \sum_{k=1}^m a_k(x)Z_k$  of  $a(x)$  admit the representation  $Z_k = \mu_k + \varepsilon \tilde{Z}_k$ , where  $\varepsilon$  is a small parameter with respect to  $\mu_k$ ,  $E[\tilde{Z}_k] = 0$ , and  $\text{Var}[\tilde{Z}_k] \sim O(1)$ ,  $k = 1, \dots, m$ . Note that the variance of  $Z_k$  is  $\varepsilon^2 \text{Var}[\tilde{Z}_k]$ .

Consider the power series expansion

$$U(x, Z) = U_0(x) + \varepsilon U_1(x) + \varepsilon^2 U_2(x) + \dots \quad (9.115)$$

for the solution of (9.53). The functions  $U_0(x), U_1(x), \dots$  satisfy the differential equations

$$\begin{aligned}
(\varepsilon^0) : \sum_{k=1}^m \mu_k \nabla \cdot (a_k(x) \nabla U_0(x)) &= -f(x) \\
(\varepsilon^1) : \sum_{k=1}^m \mu_k \nabla \cdot (a_k(x) \nabla U_1(x)) &= - \sum_{k=1}^m \tilde{Z}_k (a_k(x) \nabla U_0(x))
\end{aligned} \tag{9.116}$$

with homogeneous boundary conditions. These equations result from (9.53) with  $U(x, Z)$  in (9.115) by setting zero the coefficients of the powers of  $\varepsilon$ . Truncated power series of  $U(x, Z)$  can be used to find approximately moments and other properties of this random field. For example, the mean and correlation functions of  $U(x, Z)$  are

$$\begin{aligned}
E[U(x, Z)] &\simeq U_0(x) + O(\varepsilon^2) \\
E[U(x, Z)U(y, Z)] &\simeq U_0(x)U_0(y) + \varepsilon(U_0(x)U_1(y) + U_1(x)U_0(y)) \\
&\quad + \varepsilon^2(U_0(x)U_2(y) + U_1(x)U_1(y) + U_2(x)U_0(y)) + O(\varepsilon^3).
\end{aligned} \tag{9.117}$$

*Example 9.22* For the stochastic equation in Example 9.21, we have

$$\begin{aligned}
(\varepsilon^0) : \mu U_0''(x) &= -1, \\
(\varepsilon^1) : \mu U_1''(x) &= -\tilde{Z} U_0''(x),
\end{aligned} \tag{9.118}$$

with homogeneous boundary conditions, so that

$$U(x, Z) \simeq \frac{x - x^2}{2\mu} + \varepsilon \frac{\tilde{Z}(-x + x^2)}{2\mu^2} + O(\varepsilon^2).$$

The corresponding mean and covariance functions,

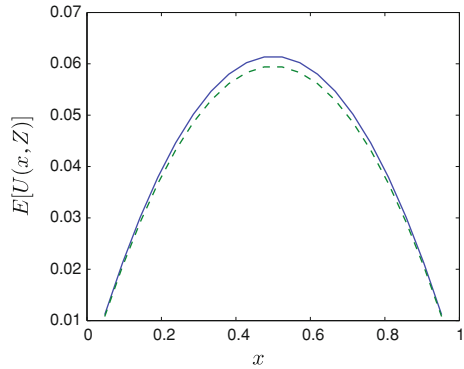
$$\begin{aligned}
E[U(x, Z)] &\simeq \frac{x - x^2}{2\mu} + O(\varepsilon) \\
\text{Cov}[U(x, Z), U(y, Z)] &\simeq \varepsilon^2 \frac{\text{Var}[\tilde{Z}]}{4\mu^4} (-x + x^2)(-y + y^2) + O(\varepsilon^3),
\end{aligned} \tag{9.119}$$

have the same expressions as those in (9.114) at the indicated accuracy.  $\diamond$

### 9.5.3 Neumann Series

Consider the differential equation  $\mathcal{L}[U(x)] = V(x)$  in (9.40) with solution  $U(x) = \mathcal{L}^{-1}[V(x)]$ . If  $\mathcal{L}^{-1}$  exists, it can be represented by a Neumann series under some conditions. Truncated Neumann series can be used to calculate properties

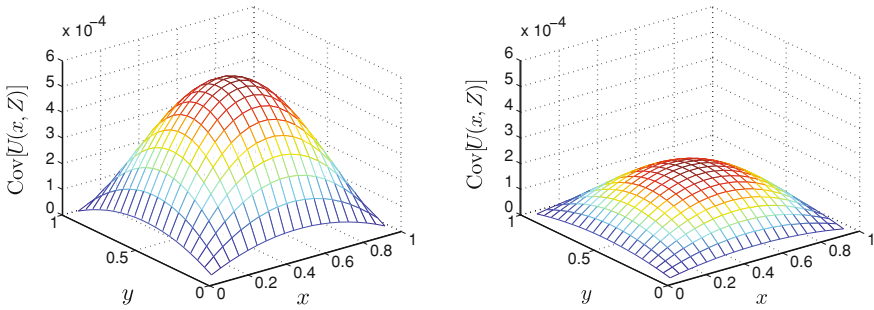
**Fig. 9.32** Approximate mean of  $U(x)$  by Neumann series (solid line) and Taylor series (dotted line) for  $\alpha = 1.2$ ,  $\beta = 3.0$ ,  $n=20$ , and  $n_{\text{NS}} = 20$



of  $U(x)$ . Neumann series solving Fredholm integral equation are discussed in [2] (Sect. 8.4.1.4). We only illustrate the construction of Neumann series for stochastic algebraic equations derived from  $\mathcal{L}[U(x)] = V(x)$  by spatial discretization.

*Example 9.23* The finite difference approximation for the stochastic differential equation in Example 9.21 has the form  $-Z(U_{i+1} - 2U_i + U_{i-1}) = h^2$  for  $i = 2, \dots, n-1$ ,  $-Z(U_2 - 2U_1) = h^2$  for  $i = 1$ , and  $-Z(-2U_n + U_{n-1}) = h^2$  for  $i = n$ , where  $h = 1/(n+1)$ ,  $U_i = U(ih)$ ,  $i = 1, \dots, n$ , and nodes 0 and  $n+1$  correspond to  $x = 0$  and  $x = 1$ . The matrix form of the finite difference formulation is  $ZAU = B$ , where  $A$  is an  $(n, n)$ -matrix with non-zero entries  $A_{i,i} = 2$  and  $A_{i,i+1} = A_{i+1,i} = -1$ ,  $B$  denotes the unit vector scaled by  $h^2$ , and  $U = (U_1, \dots, U_n)'$  is the solution vector. An alternative form of the stochastic algebraic equation  $ZAU = B$  is  $(1 + \tilde{Z}/\mu)U = (A\mu)^{-1}B$ , where  $\mu = E[Z]$  and  $\tilde{Z} = Z - \mu$ . The Neumann series for  $U$  is convergent if there exists  $0 < \gamma < 1$  such that  $\|(\tilde{Z}/\mu)x\| \leq \gamma\|x\|$  for any  $x \in \mathbb{R}^n$  (Theorem 8.19). For  $Z \sim U(\alpha, \beta)$ , the Neumann series is convergent if  $\beta - \alpha < 2\mu$ . The Neumann series representation for  $U$  is  $U = \sum_{r=0}^{\infty} U^{(r)}$ , where  $U^{(r)} = (-\tilde{Z}/\mu)U^{(r-1)}$ ,  $r = 1, 2, \dots$ , and  $U^{(0)} = (A\mu)^{-1}B$ . The mean and correlation matrices of  $U$  based on the first  $n_{\text{NS}}$  terms of the series representation of  $U$  are  $E[U(Z)] \simeq \sum_{r=1}^{n_{\text{NS}}} E[U^{(r)}]$  and  $E[U(Z)U(Z)'] \simeq \sum_{r,p=1}^{n_{\text{NS}}} E[U^{(r)}U^{(p)}]$ .

The plots in Figs. 9.32 and 9.33 are for  $Z \sim U(\alpha, \beta)$ , so that  $E[(Z - \mu)^q] = [(\beta - \mu)^{q+1} - (\alpha - \mu)^{q+1}]/(\beta - \alpha)/(q+1)$ ,  $\alpha = 1.2$ ,  $\beta = 3.0$ ,  $n = 20$  and  $n_{\text{NS}} = 20$ . Figure 9.32 shows with solid and dotted lines approximations of  $E[U(Z)]$  obtained by the Neumann and Taylor series methods. The left and right panels in Fig. 9.33 show the covariances of  $U(Z)$  by the Neumann and Taylor series methods. The first two moments of  $U(Z)$  by the Neumann series match those obtained by Monte Carlo simulation. The means of  $U(Z)$  by the Neumann and Taylor series methods are similar, but the covariances of  $U(Z)$  differ significantly. The unsatisfactory performance of the Taylor series method is caused by the approximate representation of the nonlinear mapping  $Z \mapsto U(Z)$  by the linear function of  $Z$  in (9.111).  $\diamond$



**Fig. 9.33** Approximate covariances of  $U(x, Z)$  by Neumann series (left panel) and Taylor series (right panel) for  $\alpha = 1.2$ ,  $\beta = 3.0$ ,  $n = 20$ , and  $n_{NS} = 20$

## 9.6 Exercises

**Exercise 9.1** Show that  $U^{(n)}$  in Example 9.1 is mean square continuous.

**Exercise 9.2** Complete the details of the proof in Example 9.1 showing that the weak solution of (9.5) is unique.

*Hint* Use Gronwall's inequality stating that, if  $\alpha \geq 0$  is a constant,  $\beta(t) \geq 0$  and  $\theta(t)$  are real-valued functions defined on a bounded interval  $[a, b]$  of the real line, and  $\theta(t) \leq \alpha + \int_a^t \beta(s)\theta(s) ds$ ,  $a \leq t \leq b$ , then  $\theta(t) \leq \alpha \exp \left[ \int_a^t \beta(s) ds \right]$  for  $t \in [a, b]$ .

**Exercise 9.3** Generalize considerations in Example 9.1 by assuming that the correlation of the noise process is  $E[\dot{W}(x, t)\dot{W}(y, s)] = r(x, y) \exp(-\rho|s - t|)$ ,  $\rho > 0$ , that is,  $\dot{W}(x, t)$  is a colored noise in time.

**Exercise 9.4** Consider the SPDE  $\mathcal{L}[U(x)] = W(x)$ ,  $x \in \mathbb{R}^d$ , where  $\mathcal{L} = (\Delta - \alpha^2)^p$ ,  $\alpha > 0$  is a constant,  $p \geq 1$  denotes an integer, and  $W(x)$  is a Brownian sheet, that is, a Gaussian field with mean 0 and covariance function  $E[W(x)W(y)] = \prod_{i=1}^d (x_i \wedge y_i)$ , where  $x = (x_1, \dots, x_d)$  and  $y = (y_1, \dots, y_d)$ . Find the spectral density of  $U(x)$ .

**Exercise 9.5** Develop a Monte Carlo algorithm for solving the stochastic partial differential equation in Example 9.3.

**Exercise 9.6** Check whether the bilinear form  $\mathcal{B}(U, W)$  in (9.46) defines an inner product on the set of functions  $\mathcal{W}(D)$  defined by (9.45).

**Exercise 9.7** Let  $f : [a, b] \rightarrow \mathbb{R}$  be a real-valued continuous function and  $[a, b]$  a bounded interval of the real line, so that  $f \in C[a, b]$ . Is  $\int_a^b |f(x)| dx$  a norm on  $C[a, b]$ ?

**Exercise 9.8** Let  $X$  and  $Y$  be real-valued random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Show that  $E[XY]$  defines an inner product on  $L^2(\Omega, \mathcal{F}, P)$ .

**Exercise 9.9** Find a constant  $c > 0$  satisfying the inequality in (9.61) for several distributions  $F$  of your choice. Does such a constant exist for any distribution  $F$ ?

**Exercise 9.10** Let  $a_T(x)$ ,  $x \in \mathbb{R}^2$ , be a translation random field defined by (9.85) with  $G(x) = \sum_{k=1}^n (A_k \cos(v_k \cdot x) + B_k \sin(v_k \cdot x)) / \sqrt{6}$ , where  $n = 6$ ,  $(A_k, B_k)$  are independent  $N(0, 1)$  variables,  $v^{(1)} = (1, 2)$ ,  $v^{(2)} = (2, 1)$ ,  $v^{(3)} = (2, 2)$ ,  $v^{(4)} = -v^{(1)}$ ,  $v^{(5)} = -v^{(2)}$ , and  $v^{(6)} = -v^{(3)}$ . Construct a polynomial chaos representation for  $a_T(x)$  of the type in (9.89).

**Exercise 9.11** Find the effective conductivity of an one-dimensional specimen with properties as in Example 9.12 by using SROMs.

**Exercise 9.12** Construct estimators for the mean and correlation function of  $U(x, t)$  in (9.60).

**Exercise 9.13** Calculate the second moment properties of  $K_2(x)$  in Example 9.15 and show that they coincide with those of  $K_1(x)$ .

**Exercise 9.14** Find the second moment properties of the translation model  $K(x)$  in (9.103) and compare the covariance function of this field scaled by its variance with that of  $G(x)$ .

**Exercise 9.15** Write explicitly the system of algebraic equation given by (9.101) for  $n_{PC} = 2$  and  $n = 3$ . Solve Example 9.17 for these parameters.

**Exercise 9.16** Solve the one-dimensional version of the problem in Example 9.17 by the stochastic Galerkin method.

**Exercise 9.17** Solve the one-dimensional SPDE in Example 9.18 by stochastic collocation using equally spaced collocation points in  $(0, 1)$  for  $K(x)$  given by (9.103) with  $\alpha = 0.01$ ,  $\beta = 3$ , shape parameters  $(p, q) = (1, 3)$ ,  $(1, 1)$ , and  $(3, 1)$ , and a Gaussian field  $G(x)$  with mean 0 and covariance function  $E[G(x')G(x'')] = \exp(-\rho|x' - x''|)$  for  $\rho = 0, 1$ , and 100. Assess the accuracy of the collocation solutions by Monte Carlo simulation.

## References

1. Kotelenetz P (2008) Stochastic ordinary and stochastic partial differential equations. Transition from microscopic to macroscopic equations. Springer, New York
2. Grigoriu M (2002) Stochastic calculus. Applications in science and engineering. Birkhäuser, Boston
3. Ostoja-Starzewski M (2007) Microstructural randomness and scaling in mechanics of materials. Chapman & Hall/CRC, New York
4. Sobczyk K (1985) Stochastic wave propagation. Elsevier, New York
5. Sobczyk K, Kirkner DJ (2001) Stochastic modeling of microstructures. Birkhäuser, Boston
6. Torquato S (2002) Random heterogeneous materials. Microstructure and macroscopic properties. Springer, New York

7. Babuška IM, Strouboulis T, IM The finite element method and its reliability. Clarendon Press, Oxford University Press, New York
8. Brenner SC, Scott LR (2008) The mathematical theory of finite element methods, 3rd edn. Springer, New York
9. Johnson C (1994) Numerical solution of partial differential equations by the finite element method. Cambridge University Press, New York
10. Smith GD (1985) Numerical solution of partial differential equations: finite difference methods, 3rd edn. Clarendon Press, Oxford
11. Chow P-L (2007) Stochastic partial differential equations. Chapman & Hall/CRC, Taylor & Francis Group, New York
12. Kunita H (1990) Stochastic flows and stochastic differential equations. Cambridge University Press, Cambridge
13. Milne WE, Numerical solution of differential equations. Wiley, New York
14. Lang A (2010) A Lax equivalence theorem for stochastic differential equations. *J Comput Appl Math* 234:3387–3396. doi:[10.1016/j.cam.2010.05.001](https://doi.org/10.1016/j.cam.2010.05.001)
15. Roth Ch (2002) Difference methods for stochastic partial differential equations. *ZAMM-Z Angew Math Mech* 82:821–830
16. Grecksch W, Tudor C (2000) Parabolic regularization of a first order partial differential equation. *Stoch Anal Appl* 18(3):397–416
17. Kunita H (1982) First order stochastic partial differential equations. In: Ito K (ed) *Stochastic analysis*, Kinokuniya, Tokyo
18. Reddy BD (1998) Introduction to functional analysis with applications to boundary value problems and finite elements. Springer, New York
19. Grigoriu M (2009) Reduced order models for random functions. Applications to stochastic problems. *Appl Math Model* 33(1):161–175
20. Babuška IM, Temptone R, Zouraris E (2005) Solving elliptic boundary value problems with uncertain coefficients by the finite element method: The stochastic formulation. *Comput Methods Appl Mech Eng* 194(12–14):1251–1294
21. Frauenfelder P, Schwab C, Tudor RA (2005) Finite elements for elliptic problems with stochastic coefficients. *Comput Methods Appl Mech Eng* 194:205–228
22. Grigoriu M (1995) Applied non-Gaussian processes: Examples, theory, simulation, linear random vibration, and MATLAB solutions. Prentice Hall, Englewoods Cliffs
23. Johnson N, Kotz S (1970) Distributions in statistics: Continuous univariate distributions-2. Houghton Mifflin Company, Boston
24. Adler RJ (1981) The geometry of random fields. Wiley, New York
25. Stefanou G (2000) The stochastic finite element method: Past, present and future. *Comput Methods Appl Mech Eng* 198:1031–1051
26. Grigoriu M, Kafali C (2007) Response of linear systems to stationary bandlimited non-Gaussian processes. *Probab Eng Mech* 22(4):353–361
27. Ávila da Silva CR Jr., Beck AT (2010) Bending of stochastic Kirchhoff plates on Winkler foundations via the Galerkin method and the Askey–Wiener scheme. *Probab Eng Mech* 25:172–182
28. Babuška IM, Chatzipantelidis P (2002) On solving elliptic stochastic partial differential equations. *Comput Methods Appl Mech Eng* 191:4093–4122
29. Babuška IM, Nobile F, Tempone R (2007) A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J Numer Anal* 45(3):1005–1034
30. Babuška IM, Tempone R, Zouraris E (2004) Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J Numer Anal* 42(2):800–825
31. Bressoletto Ph, Fogli M, Chauvière C (2010) A stochastic collocation method for large classes of mechanical problems with uncertain parameters. *Probab Eng Mech* 25:255–270
32. Ghanem R (1999) Ingredients for a general purpose stochastic finite elements implementation. *Comput Methods Appl Mech Eng* 168:19–34

33. Matthies HG, Keese A (2005) Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput Methods Appl Mech Eng* 194:1295–1331
34. Grigoriu M (2010) Probabilistic models for stochastic elliptic partial differential equations. *J Comput Phys* 229:8406–8429. doi:[10.1016/j.jcp.2010.07.023](https://doi.org/10.1016/j.jcp.2010.07.023)
35. Leadbetter MR, Lindgren G, Rootzén H (1983) *Extremes and related properties of random sequences and processes*. Springer, New York
36. Xiu D, *Numerical methods for stochastic computations*. Princeton University Press, Princeton
37. Foo J, Wan X, Karniadakis E (2008) The multi-element probabilistic collocation method (MEPCM): Error analysis and applications. *J Comput Phys* 227:9572–9595
38. Nobile F, Tempone R, Webster CG (2008) An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J Numer Anal* 46(5):2411–1442
39. Ganapathysubramanian B, Zabarar N (2007) Sparse grid collocation schemes for stochastic natural convection problems. *J Comput Phys* 225:652–685

# Appendix A

## Parametric Models, Quantizers, and Stochastic Reduced Order Models

### A.1 Parametric Models

Let  $X(t)$ ,  $t \in I$ , be a real/complex-valued random function with mean 0, finite variance, and correlation function  $r(s, t) = E[X(s)X(t)^*]$ . Our objective is to construct parametric models for  $X(t)$ . Arguments similar to those presented in this section can be used to construct parametric models for vector-valued random functions ([1, Chap. 6]).

#### A.1.1 Karhunen–Loève Expansion

The Karhunen–Loève expansion has been introduced in Sect. 3.6.5. It was shown that, if the correlation function  $r(s, t) = E[X(s)X(t)^*]$  is square integrable and continuous in  $I \times I$ , we have

$$X(t) = \text{l.i.m.}_{n \rightarrow \infty} \sum_{k=1}^n \lambda_k^{1/2} X_k \phi_k(t), \quad t \in I, \quad \text{where}$$

$$X_k = \lambda_k^{-1/2} \int_I X(t) \phi_k(t)^* dt, \quad E[X_k] = 0, \quad \text{and} \quad E[X_k X_l^*] = \delta_{kl} \quad (\text{A.1})$$

and  $r(t, s) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(s)^*$  (Theorem 3.22).

Parametric models can be obtained for random functions admitting Karhunen–Loève expansions by truncation. For example, the processes  $X_n(t) = \sum_{k=1}^n \lambda_k^{1/2} X_k \phi_k(t)$ ,  $n = 1, 2, \dots$ , are parametric models for  $X(t)$  obtained by truncating the expansion of  $X(t)$  in (A.1). The sequence of processes  $\{X_n(t)\}$  converges in m.s. to  $X(t)$  as  $n \rightarrow \infty$  by (A.1).

**Theorem A.1** *The mean square error of a parametric model  $X_n(t)$  for  $X(t)$  is  $E[|X(t) - X_n(t)|^2] = \sum_{k=n+1}^{\infty} \lambda_k |\phi_k(t)|^2$ .*



*Proof* The equality  $E[|X(t) - X_n(t)|^2] = \sum_{k,l=n+1}^{\infty} \lambda_k^{1/2} \lambda_l^{1/2} E[X_k X_l^*] \phi_k(t) \phi_l(t)^*$  and the property  $E[X_k X_l^*] = \delta_{kl}$  yield the stated result. The error  $E[|X(t) - X_n(t)|^2]$  approaches 0 as  $n \rightarrow \infty$  by the convergence  $X_n(t) \xrightarrow{\text{m.s.}} X(t)$ . Note also that the difference between the variances of  $X(t)$  and  $X_n(t)$  is  $\text{Var}[X(t)] - \text{Var}[X_n(t)] = \sum_{k=n+1}^{\infty} \lambda_k |\phi_k(x)|^2 \geq 0$  so that the variance of  $X_n(t)$  is smaller than that of  $X(t)$ .  $\blacktriangle$

**Theorem A.2** *If  $X(t)$  is a Gaussian function, so are its parametric models  $X_n(t)$ . If  $X(t)$  is weakly stationary,  $X_n(t)$  may not have this property.*

*Proof* The random functions  $X_n(t)$ ,  $n \geq 1$ , are Gaussian as linear forms of the Gaussian random variables  $\{X_k\}$ . That Karhunen–Loève expansions  $X_n(t)$  of weakly stationary processes  $X(t)$  can be nonstationary follows from examples showing that the variance of  $X_n(t)$  is time dependent in contrast to that of  $X(t)$  [2].  $\blacktriangle$

Karhunen–Loève expansions provide an alternative manner of specifying the probability law for Gaussian functions. However, they give no information beyond the second moment properties when dealing with non-Gaussian functions.

*Example A.1* Let  $X(t)$  and  $Y(t)$  be diffusion processes defined by the stochastic differential equations

$$\begin{aligned} dX(t) &= -\rho X(t) dt + \sqrt{2\rho} dB(t), \quad t \geq 0, \quad \text{and} \\ dY(t) &= -\rho Y(t) dt + \sqrt{\rho(3 - Y(t)^2)} dB(t), \quad t \geq 0. \end{aligned} \quad (\text{A.2})$$

The stationary solutions of these equations have the same second moment properties, that is,  $E[X(t)] = E[Y(t)] = 0$  and  $E[X(s)X(t)] = E[Y(s)Y(t)] = \exp(-\rho|s - t|)$ , but different marginal distributions, Gaussian for  $X(t)$  and uniformly distributed in  $[-\sqrt{3}, \sqrt{3}]$  for  $Y(t)$ . Yet, they share the same Karhunen–Loève expansion.  $\diamond$

*Proof* That  $E[X(t)] = E[Y(t)] = 0$  holds is left as an exercise. For  $t > s$  we have  $Y(t) = Y(s)e^{-\rho(t-s)} + \int_s^t e^{-\rho(t-u)} \sqrt{\rho(3 - Y(u)^2)} dB(u)$  so that  $E[Y(t)Y(s)] = E[Y(s)^2]e^{-\rho(t-s)}$  since  $E\left[Y(s) \int_s^t e^{-\rho(t-u)} \sqrt{\rho(3 - Y(u)^2)} dB(u)\right] = 0$ . The marginal distribution  $f_y$  of  $Y(t)$  satisfies the Fokker–Planck equation (Theorem 7.4)

$$\frac{\partial f_y(y, t)}{\partial t} = \frac{\partial(\rho y f_y(y, t))}{\partial y} + \frac{1}{2} \frac{\partial^2(\rho(3 - y^2) f_y(y, t))}{\partial y^2}$$

whose stationary solution  $f_{\text{st},y}(y) = \lim_{t \rightarrow \infty} f_y(y, t)$  is uniform in  $[-\sqrt{3}, \sqrt{3}]$ . Similar considerations show that the stationary density of  $X(t)$  is  $f_{\text{st},x}(x) = \exp(-x^2/2)/\sqrt{2\pi}$ .  $\blacktriangle$

**Definition A.1** A real-valued, weakly stationary stochastic process  $X(t)$  is said to be m.s. periodic with period  $T > 0$  if its correlation function  $r(\tau) = E[X(t)X(t + \tau)]$  is periodic with period  $T$ .

The correlation function of a m.s. periodic process  $X(t)$  with period  $T$  (Definition A.1) admits the representation  $r(\tau) = \sum_{k=-\infty}^{\infty} c_k \exp(ikv_0\tau)$ , where

$c_k \geq 0$  and  $v_0 = 2\pi/T$ . The spectral distribution  $S(v)$  of  $X(t)$  is an increasing piecewise constant function. It is common in applications to interpret  $s(v) = \sum_{k=-\infty}^{\infty} c_k \delta(v - kv_0)$  as the spectral density of  $X(t)$ .

*Example A.2* Let  $X(t)$  be a real-valued, m.s. periodic process with period  $T > 0$  and mean 0, so that the covariance function and spectral density of  $X(t)$  are  $c(\tau) = \sum_{k=-\infty}^{\infty} c_k \exp(ikv_0\tau)$  and  $s(v) = \sum_{k=-\infty}^{\infty} c_k \delta(v - kv_0)$ , where  $c_k \geq 0$  and  $v_0 = 2\pi/T$ . The Karhunen–Loève expansion of  $X(t)$  in the time interval  $I = [0, T]$  is given by (A.1) with  $\lambda_k = Tc_k$  and  $\phi_k(t) = \exp(ikv_0t)/\sqrt{T}$ ,  $k = \pm 1, \pm 2, \dots$ , and coincides with its spectral representation.  $\diamond$

*Example A.3* Let  $X(t)$ ,  $t \in \mathbb{R}$ , be a weakly stationary, real-valued stochastic process with mean 0, correlation function  $r(\tau) = E[X(t)X(t+\tau)]$ , and spectral density  $s(v)$ ,  $v \in \mathbb{R}$ . The linear operator  $\mathcal{L}[\phi(t)] = \int_{-\infty}^{\infty} r(t-s)\phi(s)ds$ ,  $t \in \mathbb{R}$ , has a continuous spectrum with eigenvalues  $\lambda = 2\pi s(v)$  and eigenfunctions  $\exp(ivt)$ ,  $v \in \mathbb{R}$ , so that its Karhunen–Loève expansion is  $X(t) = \int_{-\infty}^{\infty} e^{ivt} dZ(v)$ , where  $Z(v)$  has orthogonal increments with  $E[dZ(v)] = 0$  and  $E[|dZ(v)|^2] = s(v)dv$  [3].  $\diamond$

This example shows that Karhunen–Loève expansions developed on unbounded intervals, the entire real line in the previous example, depend on an uncountable number of random variables, so that they cannot be used to construct parametric models.

*Example A.4* Let  $X(t)$ ,  $t \in \mathbb{R}$ , be a weakly stationary process with mean 0, correlation function  $r(\tau) = E[X(t)X(t+\tau)]$ , and spectral density  $s(v)$ . Let  $\tilde{X}(t)$  be a periodic process with period  $T > 0$  such that  $\tilde{X}(t, \omega) = X(t, \omega)$ ,  $t \in (-T/2, T/2)$ . Consider the sequence of processes  $\tilde{X}_n(t)$ ,  $n = 1, 2, \dots$ , representing periodic extensions of truncated Karhunen–Loève expansions of  $X(t)$  in the time interval  $(-T/2, T/2)$ . The processes  $\tilde{X}_n(t)$  have Karhunen–Loève representations similar to those of m.s. periodic processes and converge in m.s. to  $X(t)$  in  $(-T/2, T/2)$  as  $n \rightarrow \infty$ .  $\diamond$

*Proof* Note that  $\tilde{X}(t)$  is not weakly stationary, its samples have jumps at  $t = T/2 + rT$ ,  $r \in \mathbb{Z}$ , and  $\tilde{r}(s, t) = E[\tilde{X}(s)\tilde{X}(t)]$  coincides with  $E[X(s)X(t)]$  in  $(-T/2, T/2) \times (-T/2, T/2)$  and provides a periodic extension of  $E[X(s)X(t)]$  to  $\mathbb{R}^2$ . Let  $\tilde{X}_n(t) = \sum_{k=-n}^n W_k e^{ikv_0t}$ , where  $W_k = (1/T) \int_{-T/2}^{T/2} \tilde{X}(t) e^{-ikv_0t} dt$ . Since  $E[W_k] = 0$  and

$$\tilde{r}_{kl} = E[W_k W_l^*] = (1/T^2) \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} r(t-s) e^{-i(kv_0t - lv_0s)} dt ds,$$

the correlation function  $\tilde{r}_n(t, s) = \sum_{k,l=-n}^n \tilde{r}_{kl} e^{i(kv_0t - lv_0s)}$  of  $\tilde{X}_n(t)$  represents a partial sum of the Fourier series  $\tilde{r}(t, s) = \sum_{k,l=-\infty}^{\infty} \tilde{r}_{kl} e^{i(kv_0t - lv_0s)}$  of the correlation function of  $\tilde{X}(t)$ . If  $E[X(s)X(t)]$  is continuous and has continuous partial derivatives in  $(-T/2, T/2) \times (-T/2, T/2)$ , the Fourier series of  $\tilde{r}(t, s)$  converges to  $\tilde{r}(t, s)$  in this rectangle ([4], Sect. 7.3), so that the sequence of processes  $\{\tilde{X}_n(t)\}$  converges in m.s. to  $X(t)$  as  $n \rightarrow \infty$  in  $(-T/2, T/2)$ .

The processes  $\tilde{X}_n(t)$  can be given in the form

$$\tilde{X}_n(t) = \sum_{k=-n}^n U_k \left( \sum_{l=-n}^n \beta_{lk} e^{ilv_0 t} \right), \quad (\text{A.3})$$

where  $W = (W_{-n}, \dots, W_n)$ ,  $W = \beta U$ ,  $U = (U_{-n}, \dots, U_n)$  is a random vector with uncorrelated coordinates, and  $\beta$  is a deterministic matrix that can be obtained, for example, by Cholesky's decomposition. The expressions of  $\tilde{X}_n(t)$ ,  $t \in I = (-T/2, T/2)$  resembles truncated Karhunen–Loève representations for m.s. periodic processes. ▲

### A.1.2 Spectral Representation

We have seen in Sect. 3.6.4 that m.s. continuous, weakly stationary random functions admit spectral representations of the type in (3.25) and (3.30). In contrast to Karhunen–Loève expansions that hold for both weakly stationary and nonstationary random functions, spectral representations can only be used for weakly stationary random functions. As for Karhunen–Loève expansions, spectral representations define completely the probability law for Gaussian random functions, but provide no information beyond the first two moments for non-Gaussian functions.

Parametric models based on the spectral representation theorem resemble the models used by some Monte Carlo simulation algorithms (Sect. 3.8.1). For example, let  $X(t)$  be a m.s. continuous, weakly stationary real-valued stochastic process with spectral representation

$$X(t) = \int_0^\infty [\cos(vt)dU(v) + \sin(vt)dV(v)], \quad (\text{A.4})$$

where  $E[U(v)] = E[V(v)] = 0$ ,  $E[dU(v)dV(v')] = 0$  for  $v \neq v'$ , and  $E[dU(v)^2] = E[dV(v)^2] = 2dS(v) = 2s(v)dv = g(v)dv$  for all  $v, v' \geq 0$  (Theorem 3.46). The latter equalities hold if  $S(v)$  is absolutely continuous so that it admits a spectral density  $s(v) = dS(v)/dv$  and, consequently, a one-sided spectral density  $g(v)$  (3.26).

Suppose  $X(t)$  is as in Theorem 3.18 and define the parametric models

$$X_n(t) = \sum_{k=1}^n \sigma_k [A_k \cos(v_k t) + B_k \sin(v_k t)], \quad n = 1, 2, \dots, \quad (\text{A.5})$$

where  $A_k, B_k$  are uncorrelated random variables with mean 0 and variance 1,  $0 < \bar{v} < \infty$  denotes a cutoff frequency,  $v_k = (k - 1/2)(\bar{v}/n)$ ,  $\sigma_k^2 = \int_{I_k} g(v)dv$ , and  $I_k = [v_k - \bar{v}/(2n), v_k + \bar{v}/(2n)]$  for  $k = 1, 2, \dots, n$ . We have seen that the second moment properties of  $X_n(t)$  converge to those of  $X(t)$  as the partition of  $(0, \bar{v})$  is refined, that is,  $n \rightarrow \infty$ , and  $\bar{v} \rightarrow \infty$ . Moreover, if  $X(t)$  is Gaussian, then  $X_n(t)$  is

a stationary Gaussian process for each  $n$  and becomes a version of  $X(t)$  as  $n \rightarrow \infty$  (Theorem 3.46).

**Theorem A.3** *The parametric models  $X_n(t)$  in (A.5) have mean 0 as  $X(t)$ . The discrepancies between the variances and covariance functions of  $X_n(t)$  and  $X(t)$  are  $\text{Var}[X(t)] - \text{Var}[X_n(t)] = \int_{\bar{v}}^{\infty} g(v)dv$  and*

$$c(\tau) - c_n(\tau) = \sum_{k=1}^n \int_{I_k} g(v) [\cos(v\tau) - \cos(v_k\tau)] dv + \int_{\bar{v}}^{\infty} g(v) \cos(v\tau) dv, \quad (\text{A.6})$$

for  $\bar{v}$  arbitrary, where  $c(\tau) = E[X(t)X(t + \tau)]$  and  $c_n(\tau) = E[X_n(t)X_n(t + \tau)]$ .

*Proof* The error in the variance of  $X_n(t)$  results from  $\text{Var}[X(t)] = \int_0^{\infty} g(v) dv$  and  $\text{Var}[X_n(t)] = \sum_{k=1}^n \sigma_k^2 = \int_0^{\bar{v}} g(v) dv$ . The formula in (A.6) follows from the expressions  $c(\tau) = \int_0^{\infty} g(v) \cos(v\tau) dv$  and  $c_n(\tau) = \sum_{k=1}^n \sigma_k^2 \cos(v_k\tau)$  of the covariance functions of  $X(t)$  and  $X_n(t)$ . The first term on the right side of (A.6) vanishes as  $n \rightarrow \infty$  so that  $\lim_{n \rightarrow \infty} [c(\tau) - c_n(\tau)] = \int_{\bar{v}}^{\infty} g(v) \cos(v\tau) dv$  for a fixed but arbitrary  $\bar{v}$ . The discrepancy between  $c(\tau)$  and  $c_n(\tau)$  vanishes if, in addition to mesh refining, the cutoff frequency  $\bar{v}$  is extended to infinity.  $\blacktriangle$

*Example A.5* Let  $X(t)$  be a m.s. periodic process with period  $T > 0$ , mean 0, covariance function  $c(\tau) = E[X(t + \tau)X(t)] = \sum_{k=-\infty}^{\infty} c_k \exp(ikv_0\tau)$ , and spectral density  $s(v) = \sum_{k=-\infty}^{\infty} c_k \delta(v - kv_0)$ , where  $c_k \geq 0$  and  $v_0 = 2\pi/T$ . The spectral representation of  $X(t)$  is  $X(t) = \sum_{k=-\infty}^{\infty} Z_k \exp(ikv_0t)$ , where  $\{Z_k\}$  are random variables with  $E[Z_k] = 0$  and  $E[Z_k Z_l] = c_k \delta_{kl}$ . This representation coincides with the Karhunen–Loève expansion of  $X(t)$  in Example A.2  $\diamond$

### A.1.3 Sampling Theorem

We give the classical form of the sampling theorem for deterministic functions, extend it to weakly stationary bandlimited stochastic processes, and suggest constructions of parametric models based on this theorem.

**Theorem A.4** *Let  $x : \mathbb{R} \rightarrow \mathbb{R}$  be a continuous function with Fourier transform taking non-zero values in a bounded frequency range  $[-\bar{v}, \bar{v}]$ ,  $0 < \bar{v} < \infty$ . Then*

$$x(t) = \sum_{k=-\infty}^{\infty} x(t_0 + k\tau) \alpha_k(t - t_0) \quad (\text{A.7})$$

where

$$\alpha_k(u) = \frac{\sin[\pi(u/\tau - k)]}{\pi(u/\tau - k)}, \quad (\text{A.8})$$

$\tau = \pi/\bar{v}$ , and  $t_0 \in \mathbb{R}$  is an arbitrary constant.

The theorem states that a band-limited signal  $x(t)$  is uniquely defined by its values at a countable number of arguments spaced equally at  $\tau = \pi/\bar{\nu}$ , that is, we can reconstruct  $x(t)$  from its values  $x(t_0 + k\tau)$ ,  $k = 0, \pm 1, \pm 2, \dots$ , at a countable set of times. The ratio  $1/\tau = \bar{\nu}/\pi$  is called the Nyquist sampling rate. Note also that  $\alpha_k(u) = 1$  for  $u = 0$ ,  $\alpha_k(u) = 0$  for  $u = q\tau$ ,  $q \in \mathbb{Z} \setminus \{0\}$ ,  $|\alpha_k(u)| \leq 1$  for all  $u \in \mathbb{R}$ , and  $|\alpha_k(u)| \rightarrow 0$  as  $|u| \rightarrow \infty$ .

The sampling theorem extends directly to stochastic processes whose spectral densities have a bounded support.

**Theorem A.5** *Let  $X(t)$ ,  $t \in \mathbb{R}$ , be a real-valued, weakly stationary process defined on a probability space  $(\Omega, \mathcal{F}, P)$  with mean 0, correlation function  $r(\tau) = E[X(t)X(t + \tau)]$ , and one-sided spectral density  $g(\nu)$  with support  $[0, \bar{\nu}]$ ,  $0 < \bar{\nu} < \infty$ . If  $X(t)$  has a.s. continuous samples, then*

$$X(t, \omega) = \sum_{k=-\infty}^{\infty} X(t_0 + k\tau, \omega) \alpha_k(t - t_0) \quad (\text{A.9})$$

for almost all  $\omega \in \Omega$ .

*Proof* Since  $X(t)$  can be viewed as a superposition of harmonics with random amplitudes and frequencies in the range  $[0, \bar{\nu}]$  (Theorem 3.17), its samples cannot have harmonics outside this range. Since the samples  $X(\cdot, \omega)$  of  $X(t)$  are also continuous function, Theorem A.4 can be applied sample by sample.  $\blacktriangle$

The representations in (A.7) and (A.9) involve values of  $x(t)$  and  $X(t, \omega)$  at a countably infinite set of times. We consider two approximation for the representation of  $X(t)$  that use values of  $X(t)$  at a finite number of times. The first approximation,

$$\tilde{X}_n(t) = \sum_{k=-n}^n X(t_0 + k\tau) \alpha_k(t - t_0), \quad n = 1, 2, \dots, \quad (\text{A.10})$$

is referred to as a global approximation and constitutes a truncated version of (A.9). The second approximation, referred to as a local approximation, is

$$X_n(t) = \sum_{k=n_t-n}^{n_t+n+1} X(k\tau) \alpha_k(t), \quad n = 1, 2, \dots, \quad (\text{A.11})$$

where  $n_t = [t/\tau]$  denotes the largest integer smaller than  $t/\tau$ ,  $n \geq 1$  is an integer, and  $k\tau$ ,  $k = n_t - n, \dots, n_t + n + 1$ , are time arguments, referred to as nodes. The process  $X_n(t)$  depends on the values of  $X(t)$  at  $2(n + 1)$  nodes, centered on the cell  $[n_t\tau, (n_t + 1)\tau]$  containing the current time  $t$ , and is such that  $X_n(k\tau) = X(k\tau)$  for  $k = n_t - n, \dots, n_t + n + 1$ , by the properties of  $\alpha_k$  ([5], Sect. 5.3.1.3).

**Theorem A.6** *The sequence of random variables  $\tilde{X}_n(t)$  converges in m.s. to  $X(t)$  as  $n \rightarrow \infty$  at each  $t \in \mathbb{R}$  ([6], Sect. 3.7).*

*Proof* Simple arguments show that the covariance function  $E[\tilde{X}_n(s)\tilde{X}_n(t)]$  of  $\tilde{X}_n(t)$  converges to the covariance function  $E[X(s)X(t)]$  of  $X(t)$  at any  $s, t \in \mathbb{R}$ , so that  $\tilde{X}_n(t)$  has the same second-moment properties as  $X(t)$  asymptotically as  $n \rightarrow \infty$ . If  $X(t)$  is a Gaussian process, then  $\tilde{X}_n(t)$  is Gaussian for each  $n \geq 1$  so that it becomes a version of  $X(t)$  as  $n \rightarrow \infty$ .  $\blacktriangle$

A drawback of the global representation in (A.10) is that  $n$  must increase with  $t$ , so that we need a large number of terms to represent  $X(t)$  accurately at times  $t \gg 0$ . The local approximation in (A.11) does not have this limitation.

**Theorem A.7** *The sequence of processes  $\{X_n(t)\}$  has the properties  $X_n(t) \xrightarrow{\text{m.s.}} X(t)$  at every  $t$  and  $c_n(s, t) = E[X_n(s)X_n(t)] \rightarrow c(s, t) = E[X(s)X(t)]$  at every  $(s, t)$  as  $n \rightarrow \infty$  ([5], Sect. 5.3.13).*

The representations in (A.10) and (A.11) can be constructed simply and can be used to approximate both Gaussian and non-Gaussian processes with continuous samples [3]. Note also that the processes  $\tilde{X}_n(t)$  and  $X_n(t)$  are not stationary for  $n < \infty$ , for example, the covariance function,

$$c_n(s, t) = E[X_n(s)X_n(t)] = \sum_{k=n_t-n}^{n_t+n+1} \alpha_k(t) \left( \sum_{l=n_s-n}^{n_s+n+1} c((k-l)\tau) \alpha_l(s) \right), \quad (\text{A.12})$$

of  $X_n(t)$  in (A.11) depends on  $s$  and  $t$  rather than the time lag  $|t - s|$ , and the coefficients  $X(k\tau)$  in the representation  $X_n(t)$  given by (A.11) are correlated random variables.

## A.2 Quantizers

Let  $X$  be a random element defined on a probability space  $(\Omega, \mathcal{F}, P)$  with values in a metric space  $S$  with metric  $\rho$ , that is,  $X^{-1}(\mathcal{S}) \subset \mathcal{F}$ , where  $\mathcal{S}$  denotes the Borel  $\sigma$ -field generated by the topology on  $S$  induced by metric  $\rho$ . Quantizers are simple random elements defined on  $(\Omega, \mathcal{F}, P)$  that provide approximations for  $X$  by assigning to each point in  $X(\Omega) \subset S$  a reproduction selected from a finite subset of  $S$ . If the cardinality of this subset is  $m$ , we deal with an  $m$ -quantizer.

**Definition A.2** Let  $m \geq 1$  be an integer and

$$\mathcal{V}_m = \{h : S \rightarrow S, \text{ Borel measurable and } |h| \leq m\}, \quad (\text{A.13})$$

where  $|h|$  denotes the cardinality of the quantization rule  $S \mapsto h(S)$ . Any random element  $h(X)$ ,  $h \in \mathcal{V}_m$ , is an  $m$ -quantizer for  $X$ . We are interested in an optimal  $m$ -quantizer, that is, an  $m$ -quantizer that minimizes the discrepancy between  $X$  and  $h(X)$  in some sense.

Most studies on quantizers are for  $\mathbb{R}^d$ -valued random variables  $X$  [7, 8], so that  $S = \mathbb{R}^d$ ,  $\mathcal{S} = \mathcal{B}(\mathbb{R}^d)$ . The discrepancy between  $X$  and an  $m$ -quantizer  $h(X)$ ,

$h \in \mathcal{V}_m$ , is usually measured by  $E[\|X - h(X)\|^r]$  provided  $X \in L^r(\Omega, \mathcal{F}, P)$ , where  $\|\cdot\|$  is  $\|x\| = (\sum_{i=1}^d |x_i|^r)^{1/r}$  if  $1 \leq r < \infty$  and  $\|x\| = \max_{1 \leq i \leq d} |x_i|$  if  $r = \infty$ ,  $x \in \mathbb{R}^d$ .

The error of a quantizer  $h \in \mathcal{V}_m$  with image  $h(X(\Omega)) = \{a^{(k)} \in \mathbb{R}^d, k = 1 \dots, m\}$  can be calculated from

$$E[\|X - h(X)\|^r] = \sum_{k=1}^m \int_{A_k} \|x - a^{(k)}\|^r dF(x), \quad (\text{A.14})$$

where  $A_k = h^{-1}(a^{(k)}) \in \mathcal{B}(\mathbb{R}^d)$ ,  $k = 1, \dots, m$ , and  $F$  denotes the distribution of  $X$ . The elements of  $h(X(\Omega))$ , that is, the points  $a^{(k)} \in \mathbb{R}^d$ , are referred to as centers. The optimal  $m$ -quantizer is the member of  $\mathcal{V}_m$  that has the smallest error, referred to as the minimal quantization error, so that its error is given by

$$v_{m,r}(X) = \inf_{h \in \mathcal{V}_m} E[\|X - h(X)\|^r]. \quad (\text{A.15})$$

*Example A.6* Let  $X$  denote a real-valued random variable and let  $\{A_k, k = 1, \dots, m\}$  be intervals partitioning the image  $X(\Omega)$  of  $X$ . The corresponding centers  $\{a^{(k)}\}$  of the optimal quantizer for  $r = 2$  are  $a^{(k)} = \int_{A_k} X dP / P(A_k)$  since they minimize the error  $E[\|X - h(X)\|^2] = \sum_{k=1}^m [\int_{A_k} X^2 dP - 2a^{(k)} \int_{A_k} X dP + (a^{(k)})^2 \int_{A_k} dP]$ . If  $X$  is uniformly distributed on a bounded interval, then  $\{a^{(k)}\}$  are the midpoints of the intervals  $\{A_k\}$ , so that  $\{a^{(k)}\}$  and  $\{A_k\}$  define a classical Voronoi tessellation.  $\diamond$

The following theorem shows that the quantization problem is equivalent to the  $m$ -center problem, whose objective is to find a set  $\alpha \subset \mathbb{R}^d$  of  $m$  elements in  $\mathbb{R}^d$  such that  $E[\min_{a \in \alpha} \|X - a\|^r]$  is minimized. The theorem shows that the optimal quantizer constructed on  $m$ -centers  $\{a^{(k)}\}$  is  $h = \sum_{k=1}^m a^{(k)} 1_{A_k}$ , where  $\{A_k\}$  is the Voronoi tessellation with centers  $\{a^{(k)}\}$ . The quantizer is optimal in the sense that its error is  $v_{m,r}(X)$ . An extensive and useful discussion on properties of quantizers can be found in [7] (Sect. I.4).

**Theorem A.8** *The  $m$ -quantization error  $v_{m,r}(X)$  in (A.15) can be calculated from*

$$v_{m,r}(X) = \inf_{\alpha \subset \mathbb{R}^d, |\alpha| \leq m} E\left[\min_{a \in \alpha} \|X - a\|^r\right]. \quad (\text{A.16})$$

*Proof* First note that  $Y_\alpha(\omega) = \min_{a \in \alpha} \|X(\omega) - a\|^r$ , that is, the smallest of the distances  $\|X(\omega) - a\|^r$  between a sample  $X(\omega)$  of  $X$  and the points  $a \in \alpha$ , is a sample of a positive random variable  $Y_\alpha$  depending on  $\alpha$ . The solution of the  $m$ -center problem is a set  $\alpha_{\text{opt}}$  with the property  $E[Y_{\alpha_{\text{opt}}}] \leq E[Y_\alpha]$  for all  $\alpha \subset \mathbb{R}^d$ ,  $|\alpha| \leq m$ .

Since  $h \in \mathcal{V}_m$  arbitrary defines both a partition  $\{A_k\}$  of  $X(\Omega)$  and a set of centers  $\{a^{(k)}\}$  for this partition, we have  $E[\|X - h(X)\|^r] = \sum_{k=1}^m \int_{A_k} \|x - a^{(k)}\|^r dF(x)$ . The inequality  $\int_{A_k} \|x - a^{(k)}\|^r dF(x) \geq \int_{A_k} \min_{b^{(k)} \in A_k} \|x - b^{(k)}\|^r dF(x)$  implies  $E[\|X - h(X)\|^r] \geq E[\min_{b \in \alpha} \|X - b\|^r]$ . Consider now a Voronoi tessellation  $\{A_k\}$  for  $X(\Omega)$  with centers  $\{a^{(k)}\}$ . The function  $h_V = \sum_{k=1}^m a^{(k)} 1_{A_k}$  induced by

this tessellation is a member of  $\mathcal{V}_m$  and  $E[\min_{a \in \alpha} \|X - a\|^r] = E[\|X - h_V(X)\|^r]$ . This observation and the previous inequality show that  $E[\min_{a \in \alpha} \|X - a\|^r]$  can be bounded by values of  $E[\|X - h(X)\|^r]$  for some  $h \in \mathcal{V}_m$ , so that (A.16) holds.  $\blacktriangle$

Quantizers are related to conditional expectations. Let  $X$  be a random variable with finite variance defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Let  $\{\Omega_k, k = 1, \dots, m\}$  be a measurable partition of  $\Omega$  and set  $\mathcal{G} = \sigma(\Omega_1, \dots, \Omega_m)$ . The conditional expectation  $E[X | \mathcal{G}]$  is  $\mathcal{G}$ -measurable, has the smallest m.s. error relative to  $X$ , that is,  $E[(X - E[X | \mathcal{G}])^2] \leq E[(X - Y)^2]$  for all  $\mathcal{G}$ -measurable  $Y$  with finite variance, and has the expression  $E[X | \mathcal{G}] = \sum_{k=1}^m E[X | \Omega_k] 1_{\Omega_k}$  with  $E[X | \Omega_k] = \int_{\Omega_k} X dP / P(\Omega_k)$ . Since the simple random variable  $E[X | \mathcal{G}]$  is an optimal m.s. approximation for  $X$ , it is an  $m$ -quantizer for  $r = 2$ .

*Example A.7* Let  $X \sim U(0, 1)$  and let  $A_k = [(k-1)/m, k/m)$ ,  $k = 1, \dots, m$ , be a partition of  $X(\Omega) = [0, 1]$ . The optimal  $m$ -quantizer corresponding to this partition is  $h(x) = \sum_{k=1}^m a^{(k)} 1(x \in A_k)$ , where  $a^{(k)} = \int_{A_k} x dx / \int_{A_k} dx = (k-1/2)/m$  (Example A.6), that is, a simple random variable with values  $(k-1/2)/m$  of probabilities  $1/m$ . The conditional expectation of  $X$  with respect to the  $\sigma$ -field  $\mathcal{G} = \sigma(\Omega_k = X^{-1}(A_k), k = 1, \dots, m)$  is (Sect. 2.11)

$$E[X | \mathcal{G}](\omega) = \sum_{k=1}^m 1(\omega \in \Omega_k) \frac{1}{P(\Omega_k)} \int_{\Omega_k} X dP = \sum_{k=1}^m 1(\omega \in \Omega_k) \frac{k-1/2}{m},$$

a simple random variable with the same probability law as  $h(X)$ .  $\diamond$

Since quantizers can be conditional expectations, they are likely to underestimate the uncertainty in  $X$ , as indicated by the following result. Accordingly, reliability estimates based on quantizers may be overoptimistic.

**Theorem A.9** *Let  $X$  be a real-valued random variable with finite variance that is defined on a probability space  $(\Omega, \mathcal{F}, P)$ , and let  $\mathcal{G} = \sigma(\Omega_k, k = 1, \dots, m)$  be a sub- $\sigma$ -field of  $\mathcal{F}$  corresponding to a measurable partition  $\{\Omega_k, k = 1, \dots, m\}$  of  $\Omega$ . Then*

$$E[X^2] \geq E[E[X | \mathcal{G}]^2]. \quad (\text{A.17})$$

*Proof* We have  $0 \leq E[(X - \hat{X})^2] = E[X^2] + E[\hat{X}^2] - 2E[X\hat{X}]$  with the notation  $\hat{X} = E[X | \mathcal{G}]$ . Since  $\hat{X}$  is the best m.s. estimator for  $X$ ,  $E[(X - \hat{X})Z] = 0$  for any  $\mathcal{G}$ -measurable random variable  $Z$  (Theorem 2.13). For the special case  $Z = \hat{X}$ , we have  $E[X\hat{X}] = E[\hat{X}^2]$  so that  $0 \leq E[(X - \hat{X})^2] = E[X^2] - E[\hat{X}^2]$ .  $\blacktriangle$

The construction of quantizers for random vectors can be extended to random functions provided these functions are approximated by parametric models. For example, let  $X(t)$  be a real-valued Gaussian process defined on a bounded time interval  $I \subset \mathbb{R}$  and let  $X_m(t) = \sum_{k=1}^m \lambda_k^{1/2} Z_k \varphi_k(t)$ ,  $t \in I$ , be a truncated Karhunen–Loève expansion of this process, where  $\{\lambda_k\}$  and  $\{\varphi_k(t)\}$  are the eigenvalues and the eigenfunctions of the correlation function of  $X(t)$  and  $\{Z_k\}$  are independent  $N(0, 1)$  variables. The



parametric model  $X_m(t)$  approximating  $X(t)$  depends on an  $m$ -dimensional vector  $Z = (Z_1, \dots, Z_m)$  that can be described approximately by quantizers of the type considered in this section [9, 10].

In summary, quantizers  $h(X)$  for a target random vector  $X$  are simple  $\mathbb{R}^d$ -valued random variables defined on the same probability space as  $X$  via measurable mappings  $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , the construction of quantizers involves moments of  $X$  up to an order depending on the definition of the error in (A.14), quantization and  $m$ -center problems are equivalent, and quantizers are closely related to conditional expectations for  $r = 2$ .

### A.3 Stochastic Reduced Order Models (SROMs)

There are notable similarities and differences between quantizers and stochastic reduced order models (SROMs). Both quantizers and SROMs are simple random elements that represent target random elements in some optimal sense. However, distinct optimality criteria are used to build quantizers and SROMs. Quantizers are defined on the same probability space as the random elements they approximate, while SROMs do not have to have this property. The construction of quantizers for random functions requires to develop parametric models for these functions, while the construction of SROMs does not have to use parametric models.

**Definition A.3** Let  $X$  be a random element defined on a probability space  $(\Omega, \mathcal{F}, P)$  with values in a metric space  $S$  and let  $\mathcal{S}$  be a Borel  $\sigma$ -field generated by the open balls in  $S$ . A SROM  $\tilde{X}$  for  $X$  is a simple random element with values in  $S$  whose properties are similar to those of  $X$  in the sense defined by the objective function in (A.18). If the cardinality of the range of  $\tilde{X}$  is  $m$ , we say that  $\tilde{X}$  has size  $m$ . The samples  $(\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)})$  of a SROM  $\tilde{X}$  and their probabilities  $(p_1, \dots, p_m)$  define completely  $\tilde{X}$ .

As for quantizers, the size  $m$  of SROMs is assumed to be specified since it is primarily determined by the admissible computational effort. The defining parameters  $\{(\tilde{x}^{(k)}, p_k), k = 1, \dots, m\}$  of an  $m$ -dimensional SROM  $\tilde{X}$  for a random element  $X$  are the solution of the optimization problem

$$\begin{aligned} & \min_p \{e(p)\}, \text{ with constraints} \\ & \sum_{k=1}^m p_k = 1 \text{ and } p_k \geq 0, k = 1, \dots, m, \end{aligned} \tag{A.18}$$

where  $e(p) = \sum_{u \geq 1} \alpha_u e_u(p)$  measures differences between moments, correlation, distributions, and other properties of  $X$  and  $\tilde{X}$ ,  $\alpha_u \geq 0$  are weighting factors, and  $p = (p_1, \dots, p_m)$  denotes a probability vector. Note that the samples  $\{\tilde{x}^{(k)}\}$  of  $\tilde{X}$  are numbers, vectors, or functions of time and/or space if  $X$  is a random variable, vector, or random function, respectively.

A two-step suboptimal algorithm can be used to construct SROMs. First, generate  $n_{\text{set}} \geq 1$  sets of  $m$  independent samples of  $X$ , referred to as candidate samples for  $\tilde{X}$ , and solve the optimization problem in (A.18) to find the optimal probability vector  $p^{(i)}$  and the corresponding error  $e(p^{(i)})$  for each set  $i = 1, \dots, n_{\text{set}}$  of  $m$  independent samples of  $X$ . Second, select for the defining parameters of  $\tilde{X}$  the set  $i_0$  of  $m$  independent samples and the probability vector  $p^{(i_0)}$  with the property  $e(p^{(i_0)}) = \min_{1 \leq i \leq n_{\text{set}}} \{e(p^{(i)})\}$ . Alternative algorithms for constructing SROMs can be found in [11].

SROMs are not quantizers even for random variables and vectors. Let  $h$  be a member of  $\mathcal{V}_m$  in (A.13) with centers  $\{a^{(k)}\}$ ,  $k = 1, \dots, m$ , that is,  $h$  constitutes a quantizer for a random variable/vector  $X$  defined on a probability space  $(\Omega, \mathcal{F}, P)$ . The measurable mapping  $h$  generates a measurable partition  $\{\Omega_k = X^{-1}(h^{-1}(a^{(k)}))\}$ ,  $k = 1, \dots, m$ , of  $\Omega$  so that  $\{P(\Omega_k)\}$  are the probabilities of  $\{a^{(k)}\}$ . Consider a SROM whose samples are the centers of the quantizer of  $X$ . Generally, the probabilities  $\{p_k\}$  of  $\{a^{(k)}\}$ ,  $k = 1, \dots, m$ , differ from  $\{P(\Omega_k)\}$  since they are the output of an optimization algorithm with objective function involving properties of  $X$  that differ from that in (A.14). Different probabilities  $\{p_k\}$  result for distinct definitions of the objective function  $e(p)$  in (A.18).

*Example A.8* Let  $\tilde{X}$  be a SROM with size  $m$  for an  $\mathbb{R}^d$ -valued random variable  $X$ . Errors  $e_u(p)$  in (A.18) related to distributions, moments, correlations can be defined by

$$\begin{aligned} e_1(p) &= \int_{\mathbb{R}^d} (F(x) - \tilde{F}(x))^2 w_F(x) dx, \\ e_2(p) &= \sum_{0 \leq r_1 + \dots + r_d \leq q, r_1, \dots, r_d \geq 0} (\mu(r_1, \dots, r_d) - \tilde{\mu}(r_1, \dots, r_d))^2 w_\mu(r_1, \dots, r_d), \\ e_3(p) &= \sum_{i, j = 1, \dots, d; j > i} w_r(i, j) (r(i, j) - \tilde{r}(i, j))^2, \end{aligned} \quad (\text{A.19})$$

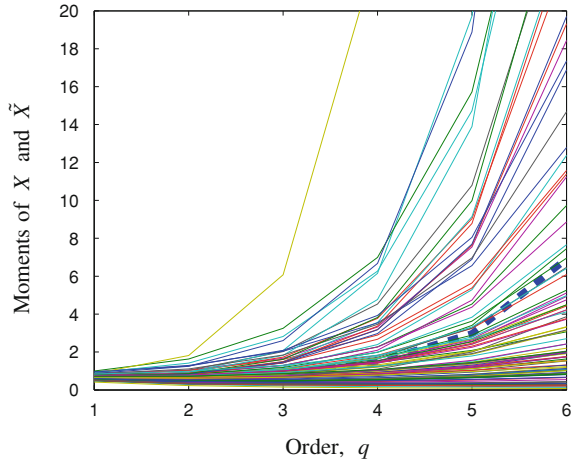
where  $w_F, w_\mu, w_r \geq 0$ , are weighting functions and  $q \geq 1$  is an integer. The distributions and the moments of  $\tilde{X}$  have the expressions

$$\begin{aligned} \tilde{F}(x) &= P(\cap_{i=1}^d \tilde{X}_i \leq x_i) = \sum_{k=1}^m p_k 1(\cap_{i=1}^d \tilde{x}_i^{(k)} \leq x_i), \\ \tilde{\mu}(r_1, \dots, r_d) &= E\left[\prod_{i=1}^d \tilde{X}_i^{r_i}\right] = \sum_{k=1}^m p_k \left(\prod_{i=1}^d (\tilde{x}_i^{(k)})^{r_i}\right), \\ \tilde{r} &= E[\tilde{X} \tilde{X}'] = \sum_{k=1}^m p_k \tilde{x}^{(k)} (\tilde{x}^{(k)})', \end{aligned} \quad (\text{A.20})$$

where  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  and  $\tilde{x}^{(k)} = (\tilde{x}_1^{(k)}, \dots, \tilde{x}_d^{(k)}) \in \mathbb{R}^d$ .

*Example A.9* Suppose  $X$  is a random variable with distribution  $F$  and moments  $\mu(q) = E[X^q] = \int x^q dF(x)$ . The stochastic reduced order model  $\tilde{X}$  for  $X$  is a

**Fig. A.1** Estimates of the first six moments of a Gamma random variable



simple random variable taking the values  $\tilde{x}^{(k)}$  with probabilities  $p_k$ ,  $k = 1, \dots, m$ , so that its distribution and moments are  $\tilde{F}(x) = \sum_{k=1}^m p_k 1(\tilde{x}^{(k)} \leq x)$  and  $\tilde{\mu}(r) = \sum_{k=1}^m p_k (\tilde{x}^{(k)})^r$ , respectively. The probability vector  $p = (p_1, \dots, p_m)$  for a specified range  $(\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)})$  of  $\tilde{X}$  minimizes the objective function

$$e(p) = \alpha_1 \int (F(x) - \tilde{F}(x))^2 dx + \alpha_2 \sum_{r=1}^q w_\mu(r) (\mu(r) - \tilde{\mu}(r))^2 \quad (\text{A.21})$$

under the constraints  $p_k \geq 0$ ,  $k = 1, \dots, m$ , and  $\sum_{k=1}^m p_k = 1$  (Example A.8). Alternative objective functions can be constructed based on entropy or other metrics.

The heavy dotted line in Fig. A.1 shows the exact first six moments of  $X$  assumed to be a Gamma random variable with density  $f(x) = x^{r-1} \eta^r \exp(-\eta x) / \Gamma(r)$ ,  $x \geq 0$ , and parameters  $r = 2$  and  $\eta = 3$ . The approximate first 6 moments of  $X$  delivered by a SROM  $\tilde{X}$  of  $X$  with  $m = 20$  obtained by a suboptimal algorithm with objective function in (A.21) with  $\alpha_1 = \alpha_2 = 1$  and  $n_{\text{set}} = 50$  are in error by less than 1%, and are indistinguishable from those of  $X$  at the figure scale. In contrast, Monte Carlo estimates based on samples of the same length as that of  $\tilde{X}$  can be inaccurate. The thin lines in Fig. A.1 are estimates of the first six moments of  $X$  corresponding to 100 sets of independent samples of  $X$  of length 20 each. Monte Carlo estimates of the first six moments of  $X$  match the accuracy of the corresponding moments of  $\tilde{X}$  with  $m = 20$  if based on approximately 50,000 independent samples of  $X$ .  $\diamond$

*Example A.10* We have seen that quantizers tend to underestimate uncertainty, which results in, for example, underestimation of higher order moments. Let  $X$  be an exponential variable with unit mean. The following table gives the exact first eight

Order	Exact	Quantizer	SROM
1	1	1	0.9994
2	2	1.9693	2.0017
3	6	5.7476	6.0004
4	24	21.64	24.03
5	120	96.69	119.49
6	720	484.74	712.24
7	5040	2622	5083
8	40320	14913	40098

moments of  $X$  and approximations of these moments based on a quantizer corresponding to the Voronoi tessellation  $I_1 = (0, (x_1 + x_2)/2)$ ,  $I_k = ((x_{k-1} + x_k)/2, (x_k + x_{k+1})/2)$ , and  $I_8 = ((x_7 + x_8)/2, \infty)$  with centers  $\{x_i\}$  at 0.1753, 0.5725, 1.0305, 1.5712, 2.2313, 3.0792, 4.2665, and 6.2665 and on a SROM with size  $m = 8$  corresponding to an objective function as in the previous example. The quantizer underestimates the moments of  $X$  and the error increases with the moment order.  $\diamond$

**Theorem A.10** *The accuracy of SROMs increases with the number  $n_{\text{set}}$  of candidate samples and with model size under some conditions. If a random element  $X$  is characterized by  $n \gg m$  independent samples, the samples of SROMs  $\tilde{X}$  with size  $m$  become equally likely as model size increases, that is,  $p_k \rightarrow 1/n$  as  $m \rightarrow n$ .*

*Proof* The performance of SROMs increases with  $n_{\text{set}}$  under the suboptimal algorithm provided the sets of samples used in  $n'_{\text{set}} > n_{\text{set}}$  trials include those used in  $n_{\text{set}}$  trials. Under full optimization, the accuracy of SROMs increases with  $m$  since the optimization space extends as  $m$  increases.

For simplicity we prove the last property for a real-valued random variable  $X$ , so that  $e(p)$  is given by (A.21), and has two components,  $e_1(p) = \int (F(x) - \tilde{F}(x))^2 dx$  and  $e_2(p) = \sum_{r=1}^q w_\mu(r) (\mu(r) - \tilde{\mu}(r))^2$ . Let  $\{x^{(i)}, i = 1, \dots, n\}$  be  $n$  independent, equally likely samples of  $X$  assumed to be sufficient to characterize  $X$ . The best SROM irrespective of size has the range  $\{x^{(i)}, i = 1, \dots, n\}$  and a probability vector  $p$  with coordinates equal to  $1/n$ , since  $e_1(p) = e_2(p) = e(p) = 0$  for this selection. Hence, the probabilities  $\{p_k\}$  for a SROM  $\tilde{X}$  with size  $m \leq n$  delivered by the optimization algorithm in (A.18) approach  $1/n$  as  $m \rightarrow n$ .  $\blacktriangle$

Generally, the discrepancy between properties of a random variable  $X$  and a SROM  $\tilde{X}$  can be viewed as consisting of two components related to the sample size  $n$  and the model size  $m$ . For example, the difference between the distributions  $F(x)$  and  $\tilde{F}(x)$  of  $X$  and  $\tilde{X}$  can be bounded by

$$(F(x) - \tilde{F}(x))^2 \leq 2[(F(x) - \hat{F}(x))^2 + (\hat{F}(x) - \tilde{F}(x))^2],$$

where  $\hat{F}(x)$  denotes the empirical distribution of  $X$  obtained from  $n$  independent samples  $\{x^{(i)}, i = 1, \dots, n\}$  of  $X$ . The term  $(F(x) - \hat{F}(x))^2$  is caused by statistical uncertainty, and can be made as small as desired by increasing the sample size since  $\hat{F}(x) \rightarrow F(x)$  in probability as  $n \rightarrow \infty$  at the continuity points of  $F$ . The term

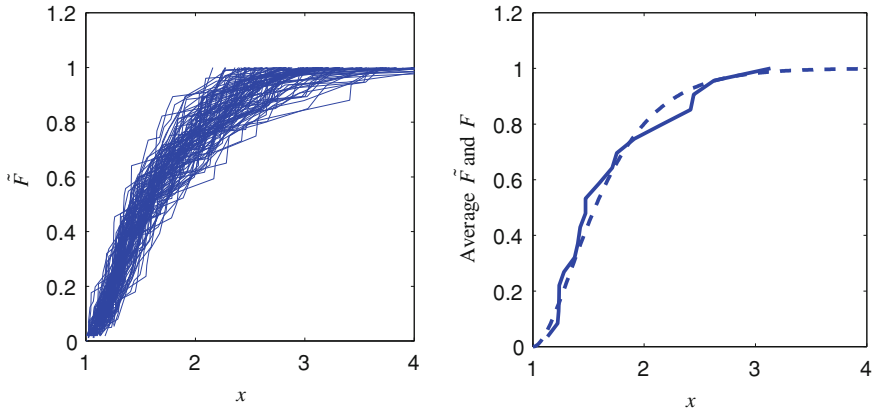
$(\hat{F}(x) - \tilde{F}(x))^2$  relates to the accuracy of  $\tilde{X}$ , and can be reduced by increasing the model size.

**Theorem A.11** *SROMs  $\tilde{X}$  can be interpreted as Borel measurable mappings of random elements  $X$  if  $X$  can be represented satisfactorily by  $n$  of independent samples.*

*Proof* Let  $(x^{(1)}, \dots, x^{(n)})$  be  $n$  independent samples of  $X$ . The construction of a Borel measurable mapping  $h : S \rightarrow S$  of the type in (A.13) is equivalent to that of a partition  $\{\mathcal{C}_k, k = 1, \dots, m\}$  of  $(x^{(1)}, \dots, x^{(n)})$  such that all  $x^{(i)}$  in  $\mathcal{C}_k$  are mapped into  $\tilde{x}^{(k)} \in S$  and  $p_k = P(X^{-1}(h^{-1}(\{\tilde{x}^{(k)}\}))) \simeq n_k/n$  for all  $k = 1, \dots, m$ , where  $n_k$  is the cardinality of  $\mathcal{C}_k$ . The partition  $\mathcal{C}_k$  of  $(x^{(1)}, \dots, x^{(n)})$  can be constructed in two steps. First, construct a preliminary partition  $\mathcal{C}'_k$  of  $(x^{(1)}, \dots, x^{(n)})$  by, for example, assigning a sample  $x^{(i)}$  to  $\mathcal{C}_k$  if it is closer to  $\tilde{x}^{(k)}$  than any other  $\tilde{x}^{(l)}$ ,  $l \neq k$ , that is,  $x^{(i)}$  is assigned to  $\mathcal{C}_k$  if  $\rho(x^{(i)}, \tilde{x}^{(k)}) < \rho(x^{(i)}, \tilde{x}^{(l)})$ ,  $l \neq k$ , where  $\rho$  is a metric in  $S$ . If  $\rho(x^{(i)}, \tilde{x}^{(k)}) = \rho(x^{(i)}, \tilde{x}^{(l)})$  for a pair  $(k, l)$ ,  $x^{(i)}$  is assigned to either  $\mathcal{C}_k$  or  $\mathcal{C}_l$ . Generally, the cardinalities  $n'_k$  of  $\mathcal{C}'_k$  do not satisfy the condition  $p_k \simeq n'_k/n$ . Second, eliminate the members of the clusters  $\mathcal{C}'_k$  with  $n'_k/n > p_k$  that are the farthest from the centers  $\tilde{x}^{(k)}$  until the modified versions  $\mathcal{C}''_k$  of these clusters satisfy the condition  $n''_k/n \simeq p_k$ , where  $n''_k$  denotes the cardinality of  $\mathcal{C}''_k$ . The members extracted from the clusters  $\mathcal{C}'_k$  with  $n'_k/n > p_k$  are assigned to the clusters  $\mathcal{C}'_l$  with  $n'_l/n < p_k$  depending on their closeness to the centers  $\tilde{x}^{(k)}$  of these clusters and the requirement  $n''_k/n \simeq p_k$ . The algorithm delivers a partition  $\{\mathcal{C}_k, k = 1, \dots, m\}$  of  $(x^{(1)}, \dots, x^{(n)})$  such that the members  $x^{(i)}$  of  $\mathcal{C}_k$  are mapped into  $\tilde{x}^{(k)}$  and the probability that  $X$  takes values in  $\mathcal{C}_k$  is  $n_k/n \simeq p_k$ . ▲

We conclude with the observation that the algorithms for constructing SROMs for random vectors extend directly to random functions. The only difference is the space in which calculations are performed. For example, a SROM  $\tilde{X}(t)$  with size  $m$  for a real-valued stochastic process  $X(t)$  consists of  $m$  samples  $(\tilde{x}_1(t), \dots, \tilde{x}_m(t))$  of  $X(t)$  and their probabilities  $(p_1, \dots, p_m)$ . Properties of  $\tilde{X}(t)$  can be obtained by elementary calculations. For example, the expectation of a functional  $\max_{0 \leq t \leq \tau} \{\tilde{X}(t)\}$  of a SROM  $\tilde{X}(t)$  is  $E[\max_{0 \leq t \leq \tau} \{\tilde{X}(t)\}] = \sum_{k=1}^m p_k (\max_{0 \leq t \leq \tau} \{\tilde{x}_k(t)\})$ .

*Example A.11* Consider the translation process  $X(t) = F^{-1} \circ \Phi(G(t))$ ,  $t \in [0, \tau]$ , where  $F$  is a Gamma distribution with shift, shape, and decay parameters  $a = 1$ ,  $k = 2$ , and  $\lambda = 3$ ,  $G(t)$  is a stationary Gaussian process with mean 0 and covariance function  $E[G(t+h)G(t)] = \exp(-|h|)$ , and  $\tau = 10$ . A SROM model  $\tilde{X}(t)$  with size  $m = 20$  has been constructed using an optimization algorithm with objective function given by (A.18) and (A.19) with  $\alpha_1 = \alpha_2 = \alpha_3 = 1$ ,  $w_F = 1$ ,  $w_\mu(\cdot) = 1/\mu(\cdot)$ ,  $w_r(\cdot) = 1/r(\cdot)$ , and  $q = 6$ . The left panel in Fig. A.2 shows marginal distributions of a SROM at 40 equally spaced times in  $[0, \tau]$ . The dotted and solid lines in the right panel are the exact and the temporal average of the marginal distributions in the left panel. The differences between temporal averages of first six moments of  $\tilde{X}$  and first six moments of  $X$  are 0.48%, 1.75%, 3.24%, 3.87%, 2.32%, and  $-2.44\%$ . ◇



**Fig.A.2** Marginal distributions of  $\tilde{X}$  and  $X$

The SROM  $\tilde{X}(t)$  is not stationary although  $X(t)$  is a stationary process. This unfavorable feature of SROMs is shared by all estimators for properties of stationary processes, and can be ameliorated by allowing the samples  $\{\tilde{x}_k(t)\}$  of  $\tilde{X}(t)$  to start at random times. Let  $\tilde{Y}(t)$  be a simple stochastic process with samples  $\{\tilde{x}_k(t + T_k)\}$  of probabilities  $\{p_k\}$ , where  $\{T_k\}$  are independent  $U(0, \bar{t})$  variables with  $\bar{t} \leq \tau$ , so that all samples are on in the time interval  $[\bar{t}, \tau - \bar{t}]$ . During this time interval, the moment of order  $r$  of  $\tilde{Y}(t)$  is

$$\begin{aligned} E[\tilde{Y}(t)^r] &= E\{E[\tilde{Y}(t)^r \mid T_1, \dots, T_m]\} = E\left\{\sum_{k=1}^m p_k (\tilde{x}_k(t + T_k))^r\right\} \\ &= \sum_{k=1}^m p_k E[(\tilde{x}_k(t + T_k))^r] = \sum_{k=1}^m p_k \frac{1}{\bar{t}} \int_0^{\bar{t}} (\tilde{x}_k(t + u))^r du, \end{aligned}$$

and constitutes a local average of the instantaneous moment  $\sum_{k=1}^m p_k (\tilde{x}_k(t))^r$  of order  $r$  of  $\tilde{X}(t)$  during  $[0, \bar{t}]$ . The temporal average in the figure is over the entire time interval  $[0, \tau]$ .  $\diamond$

## A.4 Extended Stochastic Reduced Order Models (ESROMs)

We have seen that SROMs provide efficient approximations for random elements. These models can also be used to solve stochastic equations, that is, equations with random entries. Let  $X$  be a random element defined on a probability space  $(\Omega, \mathcal{F}, P)$  that collects the random entries of a stochastic equation. The solution  $U$  of this equation is also a random element on  $(\Omega, \mathcal{F}, P)$  if the mapping  $X \mapsto U$  is measurable. Let  $\tilde{X}$  be a SROM for  $X$  with parameters  $\{\tilde{x}_k, p_k\}$ ,  $k = 1, \dots, m$ , and denote by  $\{\tilde{u}_k\}$

the solutions of this equation with  $X$  replaced by  $\{\tilde{x}_k\}$ . Properties of  $U$  can be approximated by properties of its SROM  $\tilde{U}$  with parameters  $\{\tilde{u}_k, p_k\}, k = 1, \dots, m$ . The determination of moments, distributions, and other properties of  $\tilde{U}$  involves elementary calculations.

The performance of SROM-based solutions for stochastic equations depends on the accuracy of the SROM  $\tilde{X}$  and of the representation of mapping  $X \mapsto U$ . The accuracy of  $\tilde{X}$  can be improved to a desired level by, for example, increasing the model size and/or modifying the objective function used in optimization. The mapping  $X \mapsto U$  is described by the piecewise constant function

$$\tilde{U} = \sum_{k=1}^m \tilde{u}_k 1(X \in \Gamma_k), \quad (\text{A.22})$$

where  $\{\Gamma_k\}$  is a measurable partition of the range  $\Gamma = X(\Omega)$  of  $X$  such that  $p_k = P(X^{-1}(\Gamma_k)), k = 1, \dots, m$ . This representation of mapping  $X \mapsto U$  is attractively simple but may not be sufficiently accurate.

Extended stochastic reduced order models (ESROMs) are similar to SROMs, but use piecewise linear functions, rather than piecewise constant functions as in (A.22), to approximate the mapping  $X \mapsto U$ . The supports of the local linear approximations are the cells of Voronoi tessellations centered on  $\{\tilde{x}_k\}$ . The implementation of the ESROM-based method for solving stochastic equations involves the following three steps that, for simplicity, are presented for the case in which  $X$  is an  $\mathbb{R}^d$ -valued random variable and  $U(\cdot)$  is a real-valued random function.

- *Step 1: ESROM  $\tilde{X}$  for  $X$ .* Let  $\{\tilde{x}_k\}, k = 1, \dots, m$ , be samples of  $X$ , and denote by

$$\Gamma_k = \{x \in \Gamma : \|x - \tilde{x}_k\| \leq \|x - \tilde{x}_l\|, l \neq k\}, \quad k = 1, \dots, m, \quad (\text{A.23})$$

the Voronoi tessellation with centers  $\{\tilde{x}_k\}$  in the range  $\Gamma = X(\Omega)$  of  $X$ . Any set of samples  $\{\tilde{x}_k\}$  and probabilities  $\{p_k = P(X^{-1}(\Gamma_k))\}$  defines an ESROM  $\tilde{X}$  for  $X$ . We are interested in the optimal pair  $\{\tilde{x}_k, \Gamma_k\}, k = 1, \dots, m$ , that is, the pair that minimizes the discrepancy between the probability laws of  $\tilde{X}$  and  $X$ . A suboptimal pair can be obtained by extracting subsets  $\{\tilde{x}_k\}$  of size  $m$  from a set of independent samples  $\{x_i\}, i = 1, \dots, n, n \gg m$ , of  $X$  at random; partitioning  $\{x_i\}$  in subsets  $\{\Gamma_k\}$  of the Voronoi tessellation in  $\Gamma$  given by (A.23); and calculating the discrepancy between properties of  $X$  and  $\tilde{X}$  with samples  $\{\tilde{x}_k\}$  and probabilities  $\{p_k \simeq n_k/n\}$ , where  $n_k$  gives the number of members of  $\{x_i\}$  in  $\Gamma_k$ . The ESROM  $\tilde{X}$  corresponds to the subset  $\{\tilde{x}_k\}$  of  $\{x_i\}$  with the smallest discrepancy.

- *Step 2: Approximate solutions.* Calculate the deterministic solutions  $\{\tilde{u}_k\}$  corresponding to  $\{X = \tilde{x}_k\}$  and the gradients  $\nabla \tilde{u}_k(\cdot) = (\partial \tilde{u}_k(\cdot)/\partial x_1, \dots, \partial \tilde{u}_k(\cdot)/\partial x_d), k = 1, \dots, m$ , of these solutions with respect to the coordinates of  $X$ . The gradients  $\{\nabla \tilde{u}_k\}$  can be interpreted as sensitivity factors with respect to the coordinates of  $X$ . The deterministic solutions  $\{\tilde{u}_k\}$  and  $\{\nabla \tilde{u}_k\}$  can be used to construct the local, piecewise linear approximation

$$U_L(\cdot) = \sum_{k=1}^m [\tilde{u}_k(\cdot) + \nabla \tilde{u}_k(\cdot) \cdot (X - \tilde{x}_k)] 1(X \in \Gamma_k) \quad (\text{A.24})$$

for mapping  $X \mapsto U$ , under the assumption that the mapping is sufficiently smooth. The representation in (A.24) approximates  $U(\cdot)$  in each cell  $\Gamma_k$  by a hyperplane tangent to  $X \mapsto U$  at  $(\tilde{x}_k, \tilde{u}_k)$ ,  $k = 1, \dots, m$ . Piecewise quadratic approximations can be constructed in a similar manner, but require derivatives of order two of  $U$  with respect to the coordinates of  $X$ .

- *Step 3: Solution properties.* The properties of  $U_L(\cdot)$  in (A.24) depend on the probability law of  $X$ , the samples of  $X$  that define the Voronoi tessellation  $\{\Gamma_k\}$ , and features of mapping  $X \mapsto U$ . The construction of the tessellation  $\{\Gamma_k\}$  in high dimension is a difficult task. However, it is possible to estimate properties of  $U_L$  efficiently by Monte Carlo simulation without constructing the sets  $\{\Gamma_k\}$ . Let  $\{x_i\}$ ,  $i = 1, \dots, n$ , be  $n \gg m$  independent samples of  $X$ . The members of  $\{x_i\}$  that are in  $\Gamma_k$  have the property  $\|x_i - \tilde{x}_k\| \leq \|x_i - \tilde{x}_l\|$ ,  $l \neq k$ . An algorithm can be used to identify the subsets of  $\{x_i\}$  that belong to the cells of the Voronoi tessellation in  $\Gamma = X(\Omega)$ . Denote the cardinality of these subsets by  $n_k$ . Once these subsets have been identified, properties of  $U_L$  can be estimated simply. For example, moments of order  $q \geq 1$  and marginal distributions of  $U_L$  can be estimated by

$$E[U_L(\cdot)^q] \simeq \sum_{k=1}^m \frac{n_k}{n} \left[ \frac{1}{n_k} \sum_{x_i \in \Gamma_k} (\tilde{u}_k(\cdot) + \nabla \tilde{u}_k(\cdot) \cdot (x_i - \tilde{x}_k))^q \right] \text{ and}$$

$$F_l(u, \cdot) = P(U_L(\cdot) \leq u) \simeq \sum_{k=1}^m \frac{n_k}{n} \left[ \frac{1}{n_k} \sum_{x_i \in \Gamma_k} 1(\tilde{u}_k(\cdot) + \nabla \tilde{u}_k(\cdot) \cdot (x_i - \tilde{x}_k) \leq u) \right],$$

(A.25)

from independent samples  $\{x_i\}$  of  $X$  efficiently since the functional form of  $U_L$  is known. Similar estimates can be constructed for other properties of  $U_L$ .

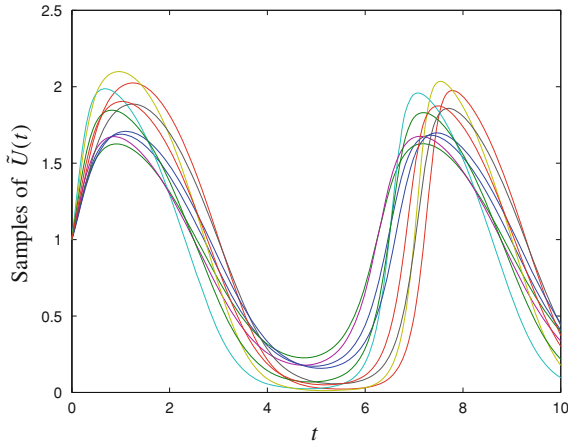
If  $X$  is a random function  $X(s)$  of time and/or space, it has to be represented by a parametric model, that is, a deterministic function  $X_P(s, Z)$  of time and/or space depending on a random vector  $Z$  (Sect. A.1). The stochastic dimension of  $X_P(s, Z)$  is finite, in contrast to that of  $X(s)$  that, generally, is not. The previous steps can be applied to solve stochastic equations depending on  $X(s)$  by replacing this random function with  $X_P(s, Z)$ .

**Example A.12** Let  $U(t)$  be the state of the dynamic system

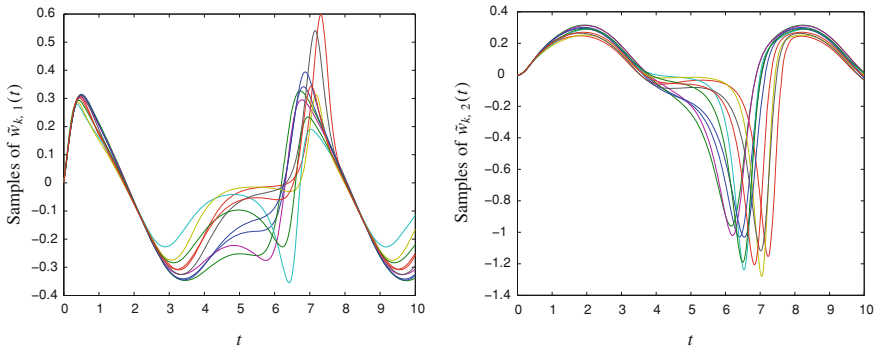
$$\dot{U}(t) = \alpha U(t) + \beta U(t)^3 + U(t) X(t), \quad t \geq 0, \quad (\text{A.26})$$

where  $X(t) = \sum_{j=1}^d Z_j \varphi_j(t)$  is a parametric input process,  $Z = (Z_1, \dots, Z_d)$  denotes an  $\mathbb{R}^d$ -valued random variable,  $\{\varphi_j(t)\}$  are specified deterministic functions, and  $\alpha, \beta$  are real constants. Let  $\{\tilde{z}_k\}$  be the samples of an ESROM model  $\tilde{Z}$  for  $Z$  and  $\{\Gamma_k\}$  the cells of a Voronoi tessellation centered on the samples of  $\tilde{Z}$ . The deterministic solutions  $\tilde{u}_k(t)$  and the coordinates  $\tilde{w}_{k,j}(t) = \partial \tilde{u}_k(t) / \partial z_j$  of the gradients of these solutions can be calculated from





**Fig.A.3** Deterministic solutions  $\tilde{u}_k(t)$

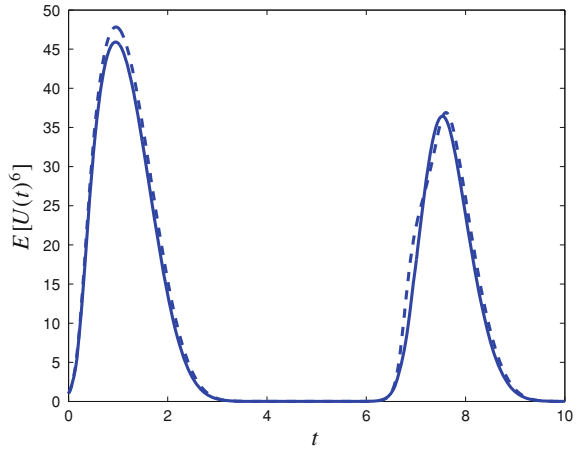


**Fig.A.4** Coordinates of gradient  $\nabla \tilde{u}_k(t) = \{\tilde{w}_{k,j}\}$  with respect to  $Z_1$  (left panel) and  $Z_2$  (right panel)

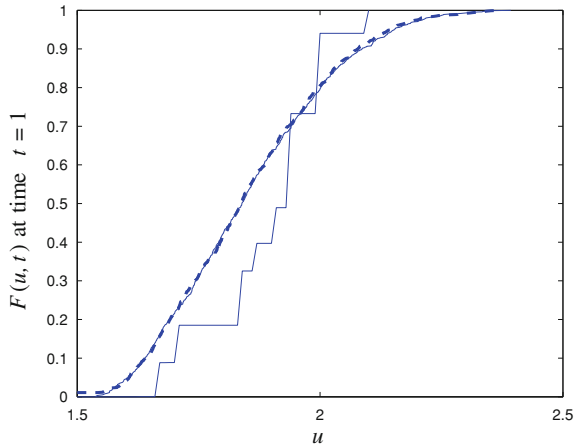
$$\begin{aligned} \dot{\tilde{u}}_k(t) &= \alpha \tilde{u}_k(t) + \beta \tilde{u}_k(t)^3 + \tilde{u}_k(t) \sum_{j=1}^d \tilde{z}_{k,j} \varphi_j(t) \text{ and} \\ \dot{\tilde{w}}_{k,j}(t) &= \left( \alpha + 3\beta \tilde{u}_k(t)^2 + \sum_{j=1}^d \tilde{z}_{k,j} \varphi_j(t) \right) \tilde{w}_{k,j}(t) + \tilde{u}_k(t) \varphi_j(t), \quad j = 1, \dots, d. \end{aligned} \tag{A.27}$$

The piecewise linear approximation  $U_L(t)$  of  $U(t)$  is given by (A.24). Following numerical results are for  $d = 2$ ,  $(Z_1, Z_2) =$  translation Beta random variables with range  $[1, 4]$ , shape parameters  $(p, q) = (1, 3)$ , and correlation coefficient  $\rho = 0.2$  between their Gaussian images,  $\alpha = 1$ ,  $\beta = -1$ , initial state  $U(0) = 1$ , time interval  $[0, 10]$ ,  $\varphi_1(t) = \cos(\nu t)$ ,  $\varphi_2(t) = \sin(\nu t)$ , and  $\nu = 1$ . The samples  $\{\tilde{z}_k\}$  of a SROM  $\tilde{Z}$  with  $m = 10$  are

**Fig. A.5** Monte Carlo estimate of  $E[U(t)^6]$  (heavy solid line) and SROM-based approximation of this moment (heavy dotted line)



**Fig. A.6** Monte Carlo estimate (thin solid line), SROM-based approximation (thin solid piecewise constant line), and ESROM-based approximation (heavy dashed line) for the marginal distribution  $F(u; t) = P(U(t) \leq u)$



$$\tilde{z}_{k,1} = 1.36; 1.39; 1.37; 2.77; 1.58; 2.30; 1.19; 1.22; 2.05; 1.71$$

$$\tilde{z}_{k,2} = 1.33; 1.00; 2.81; 1.25; 1.01; 2.55; 2.29; 1.53; 1.37; 2.03.$$

The probabilities of these samples are  $\{p_k\} = 0.0001; 0.0001; 0.2075; 0.2435; 0.0882; 0.0598; 0.0713; 0.0967; 0.1403; 0.0924$ .

Figure A.3 shows the deterministic solutions  $\tilde{u}_k(t)$  corresponding to  $Z = \tilde{z}_k$ ,  $k = 1, \dots, m$ , that is, the samples of  $\tilde{U}(t)$ . The sensitivity factors  $\tilde{w}_{k,j}$ ,  $j = 1, 2$ , are shown in Fig. A.4. The heavy solid line in Fig. A.5 is a Monte Carlo estimate of  $E[U(t)^6]$ . The dashed line is an approximation of this moment delivered by the SROM of  $U(t)$  corresponding to  $\tilde{Z}$ . The accuracy of the approximation is remarkable given that it is obtained from only  $m = 10$  deterministic solutions. The ESROM-based solution in (A.25) is indistinguishable from the Monte Carlo estimate of

$E[U(t)^6]$  at the figure scale. For this system, both SROM-based methods provide accurate approximations for  $E[U(t)^6]$ .

The thin solid and heavy dashed lines in Fig. A.6 are a Monte Carlo estimate and an ESROM-based approximation for the marginal distribution  $F(u; t) = P(U(t) \leq x)$  of  $U(t)$  at time  $t = 1$ . The piecewise constant thin solid line in the figure is an approximation of  $F(u; t)$ ,  $t = 1$ , delivered by the SROM method. The ESROM-based approximation for  $F(u; t)$  is a significant improvement over that based on the SROM method. The accurate description of the mapping  $Z \mapsto U$ , piecewise linear rather than piecewise constant representation, is the reason for the superior performance of the ESROM method.  $\diamond$

# Appendix B

## A Primer on Functional Analysis

### B.1 Metric Spaces

**Definition B.1** Let  $M$  be a set and define a function  $d : M \times M \rightarrow \mathbb{R}$  such that

$$\begin{aligned} d(x, y) &\geq 0, \quad d(x, y) = 0 \Leftrightarrow x = y && \text{(positive definite)} \\ d(x, y) &= d(y, x) && \text{(symmetry)} \\ d(x, y) &\leq d(x, z) + d(z, y) && \text{(triangle inequality)} \end{aligned} \quad (\text{B.1})$$

for all  $x, y, z \in M$ . The function  $d$  with the properties in (B.1) is said to be a metric or distance on  $M$ , and the pair  $(M, d)$  is called a metric space. A set  $M$  can be equipped with two or more metrics.

*Example B.1* Let  $M$  be an arbitrary nonempty set. Then

$$d(x, y) = \delta(x - y), \quad x, y \in M, \quad (\text{B.2})$$

is a metric on  $M$  and  $(M, d)$  is a metric space.  $\diamond$

*Example B.2* Let  $M = \mathbb{R}^q$  and define

$$\begin{aligned} d(x, y) &= \left( \sum_{i=1}^q (x_i - y_i)^2 \right)^{1/2} \\ d(x, y) &= \max_{1 \leq i \leq q} |x_i - y_i|, \end{aligned} \quad (\text{B.3})$$

where  $(x_1, \dots, x_q)$  and  $(y_1, \dots, y_q)$  denote the coordinates of  $x$  and  $y$  in  $\mathbb{R}^q$ . It is visible that these functions have the first two properties in (B.1). Since  $\max_{1 \leq i \leq q} |x_i - y_i| \leq \max_{1 \leq i \leq q} (|x_i - z_i| + |z_i - y_i|) \leq \max_{1 \leq i \leq q} |x_i - z_i| + \max_{1 \leq i \leq q} |z_i - y_i|$ , the second function  $d$  satisfies the triangle inequality, so that it is a metric on  $\mathbb{R}^q$ . That

the first function in (B.3) satisfies the triangle inequality follows from the Cauchy–Schwarz inequality given by (B.17), so that  $M = \mathbb{R}^q$  with the first distance in (B.3) is a metric space, called the  $q$ -dimensional Euclidean space. The distance  $d$  is referred to as the Euclidean metric on  $\mathbb{R}^q$ .  $\diamond$

*Example B.3* Let  $M$  consist of infinite sequences of reals  $x = (x_1, x_2, \dots)$  such that  $\sum_{i=1}^{\infty} x_i^2 < \infty$ . Then

$$\begin{aligned} d(x, y) &= \left[ \sum_{i=1}^{\infty} (x_i - y_i)^2 \right]^{1/2} \quad \text{and} \\ d(x, y) &= \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{|x_i - y_i|}{1 + |x_i - y_i|} \end{aligned} \quad (\text{B.4})$$

are metrics on  $M$ .  $\diamond$

*Proof* Consider the first function  $d(x, y)$  in (B.4). That  $d(x, y)$  satisfies the first two properties in (B.1) follows from its definition. For the third property, note that

$$\left( \sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2} \leq \left( \sum_{i=1}^n x_i^2 \right)^{1/2} + \left( \sum_{i=1}^n y_i^2 \right)^{1/2} < \infty$$

holds for any  $n \geq 1$  since  $\sum_{i=1}^n (x_i - y_i)^2 \leq \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 + 2 \sum_{i=1}^n |x_i y_i|$  and  $\sum_{i=1}^n |x_i y_i| \leq \left( \sum_{i=1}^n x_i^2 \right)^{1/2} \left( \sum_{i=1}^n y_i^2 \right)^{1/2}$  by the Cauchy–Schwarz inequality, so that  $\sum_{i=1}^{\infty} (x_i - y_i)^2 < \infty$ . The above inequality with  $x_i - z_i$  and  $z_i - y_i$  in place of  $x_i$  and  $y_i$ , respectively, gives  $d(x, y) \leq d(x, z) + d(z, y)$  by letting  $n \rightarrow \infty$ , so that  $d$  is a metric. That the second definition in (B.4) is a metric is left as an exercise.  $\blacktriangle$

*Example B.4* Let  $M = \mathbb{C}$  and set

$$d(x, y) = |x - y|, \quad x, y \in \mathbb{C}. \quad (\text{B.5})$$

This function has the first two properties in (B.1) by its definition. Since  $|a + b| \leq |a| + |b|$  holds for both real and complex numbers  $a$  and  $b$ , the function  $d$  satisfies the triangle inequality, so that it is a metric.  $\diamond$

*Example B.5* Let  $M = C[0, 1]$  denote the set of real-valued continuous functions defined on  $[0, 1]$ . The function,

$$d(x, y) = \sup\{|x(t) - y(t)| : t \in [0, 1]\}, \quad x, y \in C[0, 1], \quad (\text{B.6})$$

is a metric, so that the pair  $(C[0, 1], d)$  is a metric space. We have  $d(x, y) \geq 0$  and  $d(x, y) = d(y, x)$  by definition. The function  $d$  is strictly positive since  $d(x, y) = 0$  holds if and only if  $|x(t) - y(t)| = 0$  for every  $t \in [0, 1]$ , that is,  $x(t) = y(t)$  for every  $t \in [0, 1]$  implying  $x = y$ . The triangle inequality follows from  $|x(t) - y(t)| \leq |x(t) - z(t)| + |z(t) - y(t)| \leq d(x, z) + d(z, y)$ , which shows that  $d(x, z) + d(z, y)$  is an upper

bound for  $\{|x(t) - y(t)|, t \in [0, 1]\}$ . The set  $\{|x(t) - y(t)|, t \in [0, 1]\}$  is bounded since continuous functions on bounded intervals are bounded ([12], Theorem 16.1). Since the supremum of a set is its smallest upper bound, we have  $d(x, y) = \sup\{|x(t) - y(t)|, t \in [0, 1]\} \leq d(x, z) + d(z, y)$ , so that  $d$  satisfies the triangle inequality. The function  $d$  in (B.6) is called the sup-metric.  $\diamond$

We conclude with the observation that the collections of real-valued continuous functions defined on a compact  $K$  and of bounded real-valued functions defined on a non-empty set with the sup-metric are metric spaces. Alternative metrics on these spaces can be constructed simply. For example,  $d(x, y) = \int_K |x(t) - y(t)| dt$  is also a metric on  $C[0, 1]$  ([13], Sect. 2.2, [14], Sect. 1.1).

### B.1.1 Topology Generated by a Metric

**Definition B.2** Consider an arbitrary non-empty set  $X$ . A collection  $\mathcal{T}$  of subsets of  $X$  is a topology on  $X$  if it (1) contains the empty set  $\emptyset$  and the entire space  $X$ , (2) is closed to finite intersections, that is,  $A, B \in \mathcal{T}$  implies  $A \cap B \in \mathcal{T}$ , and (3) is closed to uncountable unions, that is,  $A_i \in \mathcal{T}, i \in I$ , implies  $\cup_{i \in I} A_i \in \mathcal{T}$ , where  $I$  is an arbitrary index set.

**Definition B.3** Let  $(M, d)$  be a metric space. The open ball or the open sphere with center  $x \in M$  and radius  $r > 0$  is the set

$$B(x, r) = \{y \in M : d(x, y) < r\}. \quad (\text{B.7})$$

**Definition B.4** A subset  $A$  of  $M$  is open if for every  $x \in A$  there exists  $r > 0$  such that  $B(x, r)$  is contained in  $A$ . The complement of an open set is a closed set. A subset of  $M$  containing a point  $x \in M$  is a neighborhood for this point if it contains an open ball centered at  $x$ .

*Example B.6* The ball  $B(x, r)$  in (B.7) is an open set since for every  $y \in B(x, r)$  we have  $B(y, r') \subseteq B(x, r)$  for  $r' < r - d(x, y)$ .  $\diamond$

*Example B.7* The ball  $B(x, r)$  in (B.7) with  $M = \mathbb{R}^2$  is a disc of radius  $r > 0$  and a square with sides  $2r$  centered at  $x \in \mathbb{R}^2$  for the first metric and the second metric in (B.3), respectively.  $\diamond$

**Theorem B.1** The collection of open sets constructed on a metric space  $(M, d)$  using open balls defines a topology  $\mathcal{T}$  on this space, referred to as the topology induced by metric  $d$ .

*Proof* Elementary arguments show that intersections of two open sets and unions of open sets are open sets ([14], Sect. 1.2, [15], Sect. III.2). Since  $M$  and  $\emptyset$  are both open and closed sets, the pair  $(M, \mathcal{T})$  is a topological space.  $\blacktriangle$

**Definition B.5** A Hausdorff, separable, or  $T_2$  space is a topological space for which distinct points have disjoint neighborhoods.

**Theorem B.2** *A metric space  $(M, d)$  is a Hausdorff space.*

*Proof* Let  $x, y \in M$ ,  $x \neq y$ , and select  $r', r'' > 0$  such that  $r' + r'' < d(x, y)$ . Then  $B(x, r')$  and  $B(y, r'')$  represent two disjoint neighborhoods of  $x$  and  $y$ , so that  $M$  with the topology induced by metric  $d$  is a Hausdorff space. ▲

**Definition B.6** Consider two topological spaces  $(X, \mathcal{T})$  and  $(X', \mathcal{T}')$ . A function  $f : X \rightarrow X'$  is said to be continuous at  $x \in X$  if for any neighborhood  $D'$  of  $f(x) \in X'$  there exists a neighborhood  $D$  of  $x \in X$  such that  $f(D) \subseteq D'$  or, equivalently,  $f^{-1}(D')$  is a neighborhood of  $x \in X$  for any neighborhood  $D'$  of  $f(x)$ . If  $f$  is continuous at every  $x \in X$ , then  $f$  is said to be continuous.

**Definition B.7** If  $M$  and  $M'$  are metric spaces with distances  $d$  and  $d'$ , respectively,  $f : M \rightarrow M'$  is continuous at  $x \in M$  if for any open ball  $B(f(x), \varepsilon)$ ,  $\varepsilon > 0$ , in  $M'$  there exists an open ball  $B(x, \delta)$ ,  $\delta > 0$ , in  $M$  such that  $f(B(x, \delta)) \subseteq B(f(x), \varepsilon)$  or, equivalently, given  $\varepsilon > 0$  there exists  $\delta > 0$  such that  $f(y) \in B(f(x), \varepsilon)$  whenever  $y \in B(x, \delta)$ .

Note that Definition B.7 is the classical  $\varepsilon$ - $\delta$  definition for continuity stating that  $f$  is continuous at  $x \in M$  if given  $\varepsilon > 0$  there exists  $\delta > 0$  such that  $d'(f(x), f(y)) < \varepsilon$  whenever  $d(x, y) < \delta$ . If  $M$  and  $M'$  are  $\mathbb{R}$ , then  $d'(f(x), f(y)) = |f(x) - f(y)|$  and  $d(x, y) = |x - y|$ .

**Theorem B.3** *Let  $(M, d)$  and  $(M', d')$  be metric spaces. A function  $f : M \rightarrow M'$  is continuous at  $x \in M$  if the convergence  $x_n \rightarrow x$  in  $(M, d)$  implies the convergence  $f(x_n) \rightarrow f(x)$  in  $(M', d')$  ([13], Sect. 2.3, [15], Sect. II.4).*

**Theorem B.4** *Let  $(M, d)$  and  $(M', d')$  be metric spaces. Suppose a function  $f : M \rightarrow M'$  has the property  $d'(f(x), f(y)) \leq cd(x, y)$ ,  $x, y \in M$ , for a fixed number  $c > 0$ . Then  $f$  is continuous.*

*Proof* The convergence  $x_n \rightarrow x$  in  $(M, d)$  means  $d(x_n, x) \rightarrow 0$  so that  $d'(f(x_n), f(x)) < cd(x_n, x) \rightarrow 0$  as  $n \rightarrow \infty$ , implying that  $f$  is continuous at  $x \in M$  by Theorem B.3. ▲

We conclude this section with two definitions involving metric spaces and mappings between these spaces.

**Definition B.8** Let  $(M, d)$  and  $(M', d')$  be metric spaces. A mapping  $f : (M, d) \rightarrow (M', d')$  is a homeomorphism if it is invertible and both  $f$  and  $f^{-1}$  are continuous functions. If such a mapping exists,  $(M, d)$  and  $(M', d')$  are said to be homeomorphic spaces.

**Definition B.9** Let  $(M, d)$  and  $(M', d')$  be metric spaces. A mapping  $f : (M, d) \rightarrow (M', d')$  is an isometry if  $d(x, y) = d'(f(x), f(y))$  for all  $x, y \in M$ . If such a mapping exists,  $(M, d)$  and  $(M', d')$  are said to be isometric spaces.

Isometric spaces are equivalent in the sense that, for example, the image  $\{f(x_n)\}$  of a sequence  $\{x_n\}$  in  $M$  converging to  $x \in M$  is convergent in  $M'$  since  $d'(f(x_n), f(x)) = d(x_n, x) \rightarrow 0$  as  $n \rightarrow \infty$ .

### B.1.2 Closed, Complete, and Compact Sets

A closed set in a topological space is the complement of an open set. This definition was used for metric spaces with topologies induced by their metrics. We give an alternative definition for closed sets, define complete and compact sets, and examine the relationship between closed, complete, and compact sets.

**Definition B.10** Let  $(M, d)$  be a metric space,  $A \subset M$ , and  $\{a_n\} \subset A^c$  a sequence converging to  $a$ . We say that  $A^c$  is closed if  $a \in A^c$ , that is,  $A^c$  contains all its accumulation points. A point  $a$  is an accumulation point for a set  $S$  if, for every  $\varepsilon > 0$ , the ball  $B(a, \varepsilon)$  contains a point  $x \in S$  distinct from  $a$ .

**Theorem B.5** Let  $A \subseteq M$  be a subset of a metric space  $(M, d)$ . Then  $A$  is open in the topology induced by metric  $d$ , that is,  $A^c$  is closed in this topology, if and only if  $A^c$  is closed according to Definition B.10, that is, the two definitions of closed sets are equivalent.

*Proof* First we show that if  $A$  is open in the topology induced by  $d$ , that is,  $A^c$  is closed in this topology, then  $A^c$  is closed by Definition B.10. Consider a sequence  $\{a_n\} \subset A^c$  converging to  $a$ , and assume  $a \in A$ . Then, there exists  $r > 0$  such that  $B(a, r) \subseteq A$  so that  $\{a_n\} \subset A^c \subset B(a, r)^c$ . Since  $d(a_n, a) \geq r$  for all  $n$ , the sequence  $\{a_n\}$  cannot converge to  $a \in A$ , so that we must have  $a \in A^c$  implying that  $A^c$  is closed according to Definition B.10.

Second we show that if  $A^c$  is closed by Definition B.10, then  $A$  is open in the topology induced by metric  $d$ . Suppose  $A$  is not open in this topology. Then, there exists  $a \in A$  such that  $B(a, 1/n) \not\subseteq A$ ,  $n = 1, 2, \dots$ , implying that there is a sequence  $a_n \in A^c$  with  $d(a, a_n) < 1/n$  that converges to  $a$ . We have constructed a convergent sequence  $\{a_n\}$  in  $A^c$  so that its limit  $a$  must be in  $A^c$  since  $A^c$  is assumed to be closed by Definition B.10. This is in contradiction with the assumption  $a \in A$ , so that  $A$  must be open in the topology induced by metric  $d$ . ▲

**Example B.8** Let  $a \in M$  be an arbitrary point of a metric space  $M$ . The set  $\{a\}$  is closed since the only sequence contained in this set is  $\{a, a, \dots\}$  and this sequence converges to  $a$ . ◇

**Example B.9** The quadrant  $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1, x \geq 0, y \geq 0\}$  is a closed set since, for example, any converging sequence  $(x_n, y_n)$  in this set satisfies the conditions  $x_n^2 + y_n^2 \leq 1$ ,  $x_n \geq 0$ , and  $y_n \geq 0$  so that its limit is also in this set. We also note that the intersection of closed sets  $\{A_i^c\}$  is itself closed since  $A = \cup_i A_i$  is open so that  $A^c = \cap_i A_i^c$  is closed. This observation can be used to show that  $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1, x \geq 0, y \geq 0\}$  is closed as an intersection of the closed sets  $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ ,  $\{(x, y) \in \mathbb{R}^2 : x \geq 0\}$ , and  $\{(x, y) \in \mathbb{R}^2 : y \geq 0\}$ . ◇

**Definition B.11** A set  $A \subset M$  of a metric space  $(M, d)$  is complete if whenever a sequence  $\{a_n\} \subset A$  has the property  $d(a_m, a_n) \rightarrow 0$  as  $m, n \rightarrow \infty$ , it is convergent



with limit in  $A$ . Sequences  $\{a_n\}$  with the property  $d(a_m, a_n) \rightarrow 0$  as  $m, n \rightarrow \infty$  are called Cauchy sequences.

**Theorem B.6** *Let  $A$  be a complete set of a metric space  $(M, d)$ . Then  $A$  is also closed.*

*Proof* Suppose  $A$  is not closed, so that we can find a sequence  $\{a_n\} \subset A$  converging to  $a \notin A$ . Since the sequence is convergent,  $d(a_m, a_n) \leq d(a_m, a) + d(a, a_n) \rightarrow 0$  as  $m, n \rightarrow \infty$  so that  $\{a_n\}$  is Cauchy. Since  $A$  is complete,  $a$  must be in  $A$  in contradiction with our assumption.  $\blacktriangle$

Arguments similar to those used to prove Theorem B.6 can be used to show that a closed subset  $A$  of a complete set  $B$  is also complete, where  $A$  and  $B$  are subsets of a metric space  $(M, d)$ . This implies that any closed subset of a complete metric space is complete.

**Theorem B.7** *The spaces  $\mathbb{R}^d$ ,  $d \geq 1$ ,  $\mathbb{C}$ , and  $C[0, 1]$  are complete metric spaces ([13], Sect. 3.4).*

*Proof* We give a partial proof of the fact that  $C[0, 1]$  with the sup-metric is complete. Let  $\{x_n\} \subset C[0, 1]$  be a Cauchy sequence, that is,  $d(x_m, x_n) \rightarrow 0$  as  $m, n \rightarrow \infty$ . Since  $0 \leq |x_m(t) - x_n(t)| \leq d(x_m, x_n)$  for each  $t \in [0, 1]$ , the numerical sequence  $\{x_n(t)\}$  is Cauchy and converges to a limit  $x(t) \in \mathbb{R}$  for each  $t \in [0, 1]$  since  $\mathbb{R}$  is complete. It remains to show that the function  $t \mapsto x(t)$  is continuous, that is,  $x \in C[0, 1]$ , and that  $d(x_n, x) \rightarrow 0$  as  $n \rightarrow \infty$ . This part of the proof can be found in [13] (Theorem 3.9).  $\blacktriangle$

**Definition B.12** A subset  $A$  of a metric space  $(M, d)$  is said to be compact if every sequence in  $A$  has a subsequence convergent to a point in  $A$ .

**Theorem B.8** *A compact set  $A$  in a metric set is also complete.*

*Proof* Let  $A$  be a compact set and let  $\{a_n\}$  be a Cauchy sequence in  $A$ , that is,  $d(a_m, a_n) \rightarrow 0$  as  $m, n \rightarrow \infty$ . Since  $A$  is compact, there exists a convergent subsequence  $\{a_{k_n}\}$  of  $\{a_n\}$  with limit  $a \in A$ . We have  $0 \leq d(a_n, a) \leq d(a_n, a_{k_n}) + d(a_{k_n}, a)$ ,  $d(a_n, a_{k_n}) \rightarrow 0$  as  $n \rightarrow \infty$  since  $\{a_n\}$  is Cauchy, and  $d(a_{k_n}, a) \rightarrow 0$  as  $n \rightarrow \infty$  since  $\{a_{k_n}\}$  is convergent. This shows that an arbitrary Cauchy sequence in  $A$  is convergent with limit in  $A$ , so that  $A$  is complete.  $\blacktriangle$

*Example B.10* The converse of Theorem B.8 is not true. For example, the real line  $\mathbb{R}$  is complete but is not compact since, for example, the sequence  $\{1, 2, \dots, n, \dots\}$  has no convergent subsequence.  $\diamond$

The statements of Theorems B.6 and B.8 can be summarized by the implications  $\text{compact} \implies \text{complete} \implies \text{closed}$ .

**Theorem B.9** *A set  $A$  is compact if and only if it is closed and bounded, that is, there exists a number  $d_0$  such that  $d(a', a'') \leq d_0$  for all  $a', a'' \in A$ .*

*Proof* We have seen that compact sets in metric spaces are also closed. Suppose a compact set  $A$  is not bounded. Then we could construct a sequence  $\{a, a_1, a_2, \dots\}$  in  $A$  such that  $d(a, a_1) \geq 1$ ,  $d(a, a_2) \geq 2$  and so on. Since the sequence cannot have a

convergent subsequence, we have a contradiction implying that  $A$  must be bounded. It remains to show that a closed and bounded set  $A$  is compact (see, for example, [13], Theorem 3.12 for  $M = \mathbb{R}^2$ ).  $\blacktriangle$

*Example B.11* A bounded interval in  $\mathbb{R}$  is closed and bounded, so that it is a compact set. However,  $\mathbb{R}$  is not compact.  $\diamond$

**Theorem B.10** *Let  $(M, d)$  and  $(M', d')$  be metric spaces and  $f : M \rightarrow M'$  be a function. Then (1)  $f$  is continuous if and only if  $f^{-1}(A)$  is closed in  $(M, d)$  whenever  $A$  is closed in  $(M', d')$ , (2) if  $M$  is compact and  $f$  is continuous, then  $f(M)$  is compact, and (3) if  $M$  is compact and  $f$  is continuous, then  $f$  is uniformly continuous.*

*Proof* For (1), suppose first that  $f$  is continuous and  $A$  is closed in  $(M', d')$ . Consider a convergent sequence  $x_n \rightarrow x$  in  $(M, d)$  such that  $\{x_n\} \subset f^{-1}(A)$ . This sequence is mapped into  $\{f(x_n)\} \subseteq A$  that converges to  $f(x)$  by continuity. Since  $A$  is closed, we have  $f(x) \in A$  so that  $x \in f^{-1}(A)$ . Conversely, assume that  $f^{-1}(A)$  is closed whenever  $A$  is closed. Consider a convergent sequence  $x_n \rightarrow x$  in  $(M, d)$  and assume that its image  $\{f(x_n)\}$  is not convergent, so that some subsequences  $\{f(x_{k_n})\}$  of  $\{f(x_n)\}$  do not approach  $f(x)$ . Note that  $A = \{x' \in M' : d(f(x_{k_n}), x') \geq \varepsilon\}$ ,  $\varepsilon > 0$ , is closed,  $\{f(x_{k_n})\} \subset A$ , and  $f^{-1}(A)$  is closed by assumption. Consider a convergent sequence  $\{x_n\} \subset f^{-1}(A)$  with limit  $x \in f^{-1}(A)$ , so that  $f(x) \in A$  implying  $d'(f(x), f(x)) \geq \varepsilon$ , a contradiction. Hence,  $\{f(x_n)\}$  must be convergent so that  $f$  is continuous ([13], Theorem 5.2).

For (2), take a sequence  $\{f(x_n)\} \subseteq M'$  that generates the sequence  $\{x_n\} \subseteq M$ , which converges to  $x \in M$  since  $M$  is compact. By continuity, we have  $f(x_n) \rightarrow f(x)$ , so that  $f(M)$  is compact ([13], Theorem 5.5).

For (3), note that  $f$  is uniformly continuous if given  $\varepsilon > 0$  there exists  $\delta > 0$  such that  $d'(f(x), f(y)) < \varepsilon$  whenever  $d(x, y) < \delta$  for all  $x, y \in M$ . Hence,  $f$  is not uniformly continuous if there exists an  $\varepsilon > 0$  such that there is no  $\delta > 0$  with the property that  $d(x, y) < \delta$  implies  $d'(f(x), f(y)) < \varepsilon$  for all  $x, y \in M$ . Fix  $\varepsilon > 0$ , set  $\delta = 1/n$ ,  $n = 1, 2, \dots$ , and let  $x_n, y_n \in M$  such that  $d(x_n, y_n) = 1/n$  and  $d'(f(x_n), f(y_n)) \geq \varepsilon$ . Since  $M$  is compact, the sequences  $\{x_n\}$  and  $\{y_n\}$  have convergent subsequences  $x_{k_n} \rightarrow x$  and  $y_{k_n} \rightarrow y$ . Since  $d(x_{k_n}, y_{k_n}) \rightarrow 0$ , we have  $x = y$  so that  $f(x_{k_n}) \rightarrow f(x)$  and  $f(y_{k_n}) \rightarrow f(y) = f(x)$  implying  $d'(f(x_{k_n}), f(y_{k_n})) \rightarrow 0$  in contradiction with  $d'(f(x_n), f(y_n)) \geq \varepsilon$ . Hence,  $f$  is uniformly continuous ([13], Theorem 5.6).  $\blacktriangle$

### B.1.3 Sequences

**Definition B.13** An infinite list of objects  $\{x_1, x_2, \dots\}$  in a set is called a sequence. A sequence  $\{x_1, x_2, \dots\}$  in a metric space  $(M, d)$  is said to be convergent and converge to a limit  $x \in M$  if  $d(x_n, x) \rightarrow 0$  as  $n \rightarrow \infty$ .

**Example B.12** Let  $M = \mathbb{R}^q$  with the Euclidean metric, and let  $x_n = (x_{n,1}, \dots, x_{n,q})$ ,  $n = 1, 2, \dots$ , be a sequence in  $M$ . Then  $x_n \rightarrow \xi = (\xi_1, \dots, \xi_q) \in \mathbb{R}^q$  if and only if  $x_{n,i} \rightarrow \xi_i$  for  $i = 1, \dots, q$ , that is, coordinate convergence. This holds since  $|x_{n,i} - \xi_i| \leq d(x^{(n)}, \xi)$ ,  $i = 1, \dots, q$ , so that the convergence  $x^{(n)} \rightarrow \xi$  implies the convergence by coordinates. Conversely, coordinate convergence, that is,  $|x_{n,i} - \xi_i| \rightarrow 0$  implies  $(x_{n,i} - \xi_i)^2 \rightarrow 0$  for all  $i = 1, \dots, q$  so that  $d(x_n, \xi) \rightarrow 0$  as  $n \rightarrow \infty$ .  $\diamond$

**Example B.13** A sequence may or may not be convergent depending on the metric used to assess this property. For example, consider the sequence  $x^{(n)} = (2 + 1/2^n, 1/2^n) \in \mathbb{R}^2$ ,  $n = 1, 2, \dots$ , that converges to  $(2, 0)$  with respect to the metrics in (B.3). However, the sequence diverges under the metric in (B.2) since  $d(x^{(n)}, (2, 0)) = 1$  for all  $n \geq 1$  ([13], Sect. 2.3).  $\diamond$

**Example B.14** Let  $M = C[1, 2]$  be a metric space with the sup-metric, and let  $x_n(t) = nt/(n+t)$  be a sequence in this space. The distance between  $x_n$  and  $x(t) = t$ , a member of  $C[1, 2]$ , is  $d(x_n, x) = t^2/(n+t) \leq t^2/n \leq 4/n \rightarrow 0$  so that  $x_n \rightarrow x$ . However, coordinate convergence, that is, the convergence  $x_n(t) \rightarrow x(t)$  for each value of  $t \in [1, 2]$ , does not necessarily imply  $x_n \rightarrow x$ . For example, the sequence  $x_n(t) = t^n$ ,  $n = 1, 2, \dots$ , in  $C[0, 1]$  converges coordinate-by-coordinate to  $x(t) = \delta(t-1)$ . However,  $x_n$  does not converge to  $x$  in  $C[0, 1]$  since  $x \notin C[0, 1]$ .  $\diamond$

**Theorem B.11** Let  $\{x_n\}$  be a convergent sequence in a metric space  $(M, d)$ . Then (1)  $x_n$  has a unique limit, (2) any subsequence of  $x_n$  is convergent and converges to the same limit, and (3)  $x_n$  is a Cauchy sequence, that is,  $d(x_m, x_n) \rightarrow 0$  as  $m, n \rightarrow \infty$ .

*Proof* Suppose  $x_n$  has two limits,  $x$  and  $y$ . Since  $d(x, y) \leq d(x, x_n) + d(x_n, y)$  and  $d(x, x_n), d(y, x_n) \rightarrow 0$  as  $n \rightarrow \infty$ , we have  $d(x, y) = 0$ , that is,  $x = y$ . Let  $x_{k_1}, x_{k_2}, \dots$ ,  $k_1 < k_2 < \dots$ , be a subsequence of  $x_n$ . We have  $d(x_{k_n}, x) \rightarrow 0$  as  $n \rightarrow \infty$  since  $d(x_n, x) \rightarrow 0$  by assumption and  $\{x_{k_n}\} \subset \{x_n\}$ . That  $x_n$  is Cauchy follows from the triangle inequality.  $\blacktriangle$

### B.1.4 Contraction

**Definition B.14** Let  $(M, d)$  be a metric space. A mapping  $T : M \rightarrow M$  is said to be a contraction if there exists a real number  $\lambda \in [0, 1)$ , called contraction factor, such that  $d(Tx, Ty) \leq \lambda d(x, y)$  for all  $x, y \in M$ , where  $Tx$  is a short hand notation for  $T(x)$ .

**Definition B.15** A fixed point for  $T : M \rightarrow M$  is a point  $x \in M$  such that  $Tx = x$ .

**Theorem B.12** Contractions are continuous functions and  $d(T^n x, T^n y) \leq \lambda^n d(x, y)$ , where  $T^n$  is an abbreviation for  $T \circ T^{n-1}$ ,  $n \geq 2$ .

*Proof* Let  $x_n \rightarrow x$  be a convergent sequence, so that  $d(x_n, x) \rightarrow 0$  as  $n \rightarrow \infty$ . Hence,  $d(Tx_n, Tx) \leq \lambda d(x_n, x) \leq d(x_n, x)$  also converges to 0 as  $n \rightarrow \infty$ , so that

$T$  is continuous. The inequality  $d(T^n x, T^n y) \leq \lambda^n d(x, y)$  results from the defining property of contraction. ▲

**Theorem B.13** (The Banach fixed point theorem) *Any contraction  $T : M \rightarrow M$  in a complete metric space  $(M, d)$  has a unique fixed point  $x_0 \in M$ . If  $x_0$  is a fixed point, then  $T^n x_0 \rightarrow x_0$  as  $n \rightarrow \infty$  ([14], Theorem 1.3.15).*

## B.2 Linear or Vector Spaces

**Definition B.16** A set  $V$  is said to be a linear or vector space if it is closed under finite vector addition and scalar multiplication, that is, for any  $x, y, z \in V$  and scalars  $\alpha, \beta \in \mathbb{F}$ , we have

$$\begin{aligned}
 x + y &= y + x && \text{(commutativity)} \\
 (x + y) + z &= x + (y + z) && \text{(addition associativity)} \\
 0 + x &= x + 0 = x && \text{(addition identity)} \\
 x + (-x) &= 0 && \text{(existence of additive inverse)} \\
 \alpha(\beta x) &= (\alpha\beta)x && \text{(associativity of scalar multiplication)} \\
 (\alpha + \beta)x &= \alpha x + \beta x && \text{(distributivity of scalar sums)} \\
 \alpha(x + y) &= \alpha x + \alpha y && \text{(distributivity of vector sum)} \\
 1x &= x && \text{(scalar multiplication identity).} \quad (\text{B.8})
 \end{aligned}$$

It is sufficient for our discussion to assume that the field  $\mathbb{F}$  coincides with  $\mathbb{R}$  or  $\mathbb{C}$ . The linear spaces corresponding to  $\mathbb{R}$  and  $\mathbb{C}$  are called real and complex, respectively.

**Definition B.17** A subset  $V_0$  of  $V$  is called a linear subspace of  $V$  if  $x + y \in V_0$  and  $\alpha x \in V_0$  for  $x, y \in V_0$  and arbitrary scalar  $\alpha$ . Note that a linear subspace is a linear space and that the intersection of linear subspaces including an arbitrary set in  $V$  is a linear subspace.

*Example B.15* The collection of vectors in  $\mathbb{R}^d$ ,  $d \geq 1$ , with the usual addition and scalar multiplication is a linear space. The vectors contained in an arbitrary plane define a linear subspace of  $\mathbb{R}^3$ . ◇

*Example B.16* The set  $C[0, 1]$  of real-valued continuous function with pointwise addition and multiplication, that is,  $(x + y)(t) = x(t) + y(t)$  and  $(\alpha x)(t) = \alpha x(t)$ ,  $x, y \in C[0, 1]$ ,  $t \in [0, 1]$ , is a linear space. These operations are similar to those in  $\mathbb{R}^d$  if we view  $x(t)$ ,  $t \in [0, 1]$ , as the coordinates of the infinitely dimensional vector defining  $x$ . Note also that the set of real-valued continuous functions  $x$  defined on  $[0, 1]$  with the property  $x(0) = x(1) = 0$  defines a linear subset of  $C[0, 1]$ . Additional examples can be found in [16] (Sect. 4.2). ◇

**Definition B.18** A finite subset  $B = (x_1, \dots, x_n)$  of  $V$  is linearly independent if  $\sum_{i=1}^n \alpha_i x_i = 0$ ,  $\alpha_i \in \mathbb{F}$ , implies  $\alpha_1 = \dots = \alpha_n = 0$ . If every  $x \in V$  has the unique representation  $x = \sum_{i=1}^n \alpha_i x_i$ ,  $\alpha_i \in \mathbb{F}$ , we say that  $B$  is a basis for  $V$  and that  $V$  is a

finite dimensional space with dimension  $n$ . The scalars  $\{\alpha_i\}$  are called the coordinates of  $x$  in basis  $B$ .

*Example B.17* The  $n$ -dimensional vectors  $(1, 0, 0, \dots)$ ,  $(0, 1, 0, \dots)$ ,  $\dots$  define a basis for the Euclidean space  $\mathbb{R}^n$ .  $\diamond$

**Theorem B.14** *A finite dimensional vector space  $V$  may have many bases but they have the same number of members, and this number defines the dimension of  $V$ .*

*Proof* Let  $B = (x_1, \dots, x_n)$  and  $B' = (y_1, \dots, y_m)$  be bases in  $V$ . Suppose  $m > n$  and  $(y_1, \dots, y_n)$  are linearly independent. Since  $y_j \in V$ , we have  $y_j = \sum_{i=1}^n a_{j,i} x_i$ ,  $j = 1, \dots, m$ . Let  $b, c \in \mathbb{F}$  be such that  $by_k + cy_l = \sum_{i=1}^n (ba_{k,i} + ca_{l,i}) x_i = 0$ . Since  $(x_1, \dots, x_n)$  are linearly independent, we have  $ba_{k,i} + ca_{l,i} = 0$  or  $a_{l,i} = -(b/c)a_{k,i}$ ,  $i = 1, \dots, n$ , so that the vectors  $y_l$ ,  $k \leq n < l$ , are linearly dependent on  $(y_1, \dots, y_n)$ , which implies that  $B'$  has only  $n$  linearly independent vectors. Similar arguments show that  $B'$  cannot have fewer than  $n$  linearly independent vectors.  $\blacktriangle$

*Example B.18* Let  $B = (x_1, \dots, x_n)$  be a basis for an  $n$ -dimensional vector space  $V$ . Let  $x'_i = \sum_{j=1}^n c_{ij} x_j$ ,  $i = 1, \dots, n$ , and denote by  $c = \{c_{ij}\}$  the matrix defining the mapping  $(x_1, \dots, x_n) \mapsto (x'_1, \dots, x'_n)$ . If  $c$  is not singular, then  $B' = (x'_1, \dots, x'_n)$  is a basis for  $V$ . This follows from the observation that  $B'$  is a basis in  $V$  if its members are linearly independent, that is,  $\sum_{i=1}^n \alpha_i x'_i = 0$  holds only if  $\alpha_i = 0$ ,  $i = 1, \dots, n$ , or  $\sum_{j=1}^n (\sum_{i=1}^n c_{ij} \alpha_i) x_j = 0$  implying  $\sum_{i=1}^n c_{ij} \alpha_i = 0$ ,  $j = 1, \dots, n$ , since  $(x_1, \dots, x_n)$  are linearly independent. This system of equation has the solution  $\alpha_1 = \dots = \alpha_n = 0$  if the determinant of  $c$  is not 0.  $\diamond$

**Theorem B.15** *The linear mapping  $f : V \rightarrow \mathbb{F}^n$  defined by  $f(v) = (\alpha_1, \dots, \alpha_n)$  is an isomorphism of  $V$  onto  $\mathbb{F}^n$ , where  $(\alpha_1, \dots, \alpha_n)$  are the coordinates of  $x \in V$  in a basis  $B = (x_1, \dots, x_n)$ .*

*Proof* The function  $f : V \rightarrow \mathbb{F}^n$  is linear with inverse  $f^{-1} : \mathbb{F}^n \rightarrow V$  defined by  $f^{-1}(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i x_i \in V$ . Since  $f$  is a bijection or a one-to-one correspondence, that is, whenever  $f(x) = f(y)$  then  $x = y$ , and an onto function, that is, all members of  $\mathbb{F}^n$  are used,  $f$  is an isomorphism.  $\blacktriangle$

### B.3 Normed Linear Spaces

**Definition B.19** A function  $\|\cdot\| : V \rightarrow \mathbb{R}$  defined on a linear space  $V$  is a norm if

$$\begin{aligned} \|x\| > 0, x \neq 0, \quad \|x\| = 0 &\Leftrightarrow x = 0 && \text{(positive definite)} \\ \|\alpha x\| = |\alpha| \|x\| &&& \text{(linearity)} \\ \|x + y\| \leq \|x\| + \|y\| &&& \text{(triangle inequality)} \end{aligned} \quad (\text{B.9})$$

for all  $x, y \in V$  and any scalar  $\alpha$ . The definition requires the linear space structure so that, for example,  $x + y, \alpha x \in V$  for  $x, y \in V$  and scalar  $\alpha$ . The pair  $(V, \|\cdot\|)$  is

called a normed linear space or normed vector space. If  $x \mapsto \|x\|$  does not satisfy the condition  $\|x\| = 0 \Leftrightarrow x = 0$ ,  $\|\cdot\|$  is said to be a seminorm.

**Example B.19** The norm of a vector can be interpreted as its length. For example, the length of the vectors  $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ , referred to as the Euclidean norm, is  $\|x\| = (x_1^2 + x_2^2 + x_3^2)^{1/2}$ .  $\diamond$

**Theorem B.16** All norms on a finite dimensional vector space  $V$  are equivalent in the sense that a sequence converges in a norm if and only if converges in any other norm.

*Proof* Let  $(e_1, \dots, e_n)$  be a basis in  $V$ , so that  $x \in V$  admits the representation  $x = \sum_{i=1}^n \alpha_i e_i$ ,  $\alpha_i \in \mathbb{C}$ . For an arbitrary norm  $\|\cdot\|$  on  $V$ , we have  $\|x\| = \|\sum_{i=1}^n \alpha_i e_i\| \leq (\sum_{i=1}^n \|e_i\|^2)^{1/2} (\sum_{i=1}^n |\alpha_i|^2)^{1/2} = c \|x\|_1$  by the Cauchy–Schwarz inequality with the notation  $c = (\sum_{i=1}^n \|e_i\|^2)^{1/2} > 0$  and  $\|x\|_1 = (\sum_{i=1}^n |\alpha_i|^2)^{1/2}$  denotes another norm. Let  $h(\beta) = \|\sum_{i=1}^n \beta_i e_i\|$  be a function defined on  $S = \{\beta = (\beta_1, \dots, \beta_n) \in \mathbb{C}^n : \|\beta\| = 1\}$ . Since  $h$  is continuous and  $S$  is compact, there exists  $\xi \in S$  such that  $h(\xi) = \min_{\beta \in S} h(\beta)$ . Moreover,  $h(\xi) > 0$  since  $(e_1, \dots, e_n)$  are linearly independent vectors so that  $\|x\| = \|\alpha\| h(\alpha/\|\alpha\|) \geq h(\xi) \|x\|_1$ . We have  $h(\xi) \|x\|_1 \leq \|x\| \leq c \|x\|_1$  so that convergence in norm  $\|\cdot\|_1$  implies convergence in norm  $\|\cdot\|$ , which proves the theorem since  $\|\cdot\|$  and  $\|\cdot\|_1$  are arbitrary.  $\blacktriangle$

**Definition B.20** A Banach space is a normed linear space  $(V, \|\cdot\|)$  that is complete. Note that any normed linear space with finite dimension is a Banach space ([16], Theorem 5.10.2).

**Example B.20** Let  $l^p$ ,  $1 \leq p < \infty$ , be the set of all sequences  $x = (x_1, x_2, \dots)$  of scalars such that  $\sum_{k=1}^{\infty} |x_k|^p < \infty$ . The set  $l^p$  with operations similar to those in  $\mathbb{R}^d$  is a linear space, and with the  $p$ -norm

$$\|x\|_p = \left( \sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} \quad (\text{B.10})$$

is a Banach space. Similarly, the space  $l^\infty$  of all bounded sequences  $x = (x_1, x_2, \dots)$  with operation as in  $l^p$  is a linear space that with the norm

$$\|x\|_\infty = \sup\{|x_k| : 1 \leq k < \infty\} \quad (\text{B.11})$$

is a Banach space. The norms in (B.10) and (B.11) defined for infinite-dimensional vector spaces can be also used for vector spaces with finite dimension.  $\diamond$

**Example B.21** Let  $L^p[0, 1]$ ,  $1 \leq p < \infty$ , consists of real-valued functions defined on  $[0, 1]$  such that  $\int_0^1 |x(t)|^p dt < \infty$ . Then  $L^p[0, 1]$  with pointwise addition and scalar multiplication is a linear space and, with the norm

$$\|x\|_p = \left( \int_0^1 |x(t)|^p dt \right)^{1/p}, \quad (\text{B.12})$$

is a Banach space. It is possible to have  $\|x - y\| = 0$  although  $x$  and  $y$  differ on a set of zero Lebesgue measure. To remove this ambiguity, we will not distinguish between members of  $L^p[0, 1]$  that are equal almost everywhere (a.e.), that is, members  $x, y \in L^p[0, 1]$  such that  $\|x - y\| = 0$ . A broad range of examples of normed linear spaces can be found in [16] (Sect. 5.3)  $\diamond$

**Example B.22** The space of real-valued continuous functions  $C[0, 1]$  with the norm  $\|x\| = \|x\|_\infty = \sup\{|x(t)| : t \in [0, 1]\}$  is a Banach space, that is, any Cauchy sequence  $\{x_n\} \subset C[0, 1]$  is convergent and converges to a member of  $C[0, 1]$  ([13], Theorem 3.9). On the other hand,  $C[0, 1]$  with the norm  $\|x\| = \int_0^1 |x(t)| dt$  is not complete. For example, the sequence of continuous functions  $x_n(t) = 1(0 \leq t \leq 1/2) + 1(1/2 \leq t \leq 1/2 + 1/n)(1 - n(t - 1/2))$ ,  $n = 1, 2, \dots$ , is Cauchy since  $\|x_{n+k} - x_n\| \leq \int_{1/2}^{1/2+1/n} x_n(t) dt \sim O(n^{-1})$  but converges to  $x(t) = 1(0 \leq t \leq 1/2)$ , that is not in  $C[0, 1]$ .  $\diamond$

In linear spaces, only finite sums of vectors are defined. This restriction can be removed when dealing with normed linear spaces  $(V, \|\cdot\|)$ , so that infinite sum  $\sum_{i=1}^\infty x_i$ ,  $x_i \in V$ , have meaning.

**Definition B.21** Let  $(V, \|\cdot\|)$  be a normed linear space,  $\sum_{i=1}^\infty x_i$ ,  $x_i \in V$ , an infinite sum, and  $s_n = \sum_{i=1}^n x_i$ ,  $n = 1, 2, \dots$ , partial sums of  $\sum_{i=1}^\infty x_i$ . If  $\{s_n\}$  is a convergent sequence, that is, there exists  $s \in V$  such that  $\|s_n - s\| \rightarrow 0$  as  $n \rightarrow \infty$ , then the notation  $\sum_{i=1}^\infty x_i$  has meaning. We say that the sum  $\sum_{i=1}^\infty x_i$  is convergent and equal  $s = \lim_{n \rightarrow \infty} s_n$ .

**Theorem B.17** If  $\{s_n\}$  is a convergent sequence, then  $\|x_n\| \rightarrow 0$  as  $n \rightarrow \infty$ .

*Proof* Since  $\{s_n\}$  is convergent, we have  $\|s_m - s_n\| \leq \|s_m - s\| + \|s - s_n\| \rightarrow 0$  as  $m, n \rightarrow \infty$ , which implies  $\|x_n\| \rightarrow 0$ ,  $n \rightarrow \infty$ , by taking  $m = n + 1$ .  $\blacktriangle$

**Definition B.22** Let  $V'$  and  $V''$  be linear subspaces of a normed linear space  $V$  whose only common point is  $0 \in V$ . If for every  $x \in V$  there exists  $x' \in V'$  and  $x'' \in V''$  such that  $x = x' + x''$ , we say that  $V$  is the direct sum of  $V'$  and  $V''$  and use the notation  $V = V' \oplus V''$ .

**Theorem B.18** The decomposition  $x = x' + x''$  is unique.

*Proof* If  $x$  has another representation  $x = y' + y''$ ,  $y' \in V'$  and  $y'' \in V''$ , we have  $x' - y' = x'' - y''$  with  $x' - y' \in V'$  and  $x'' - y'' \in V''$ . Since  $0$  is the only common element of  $V'$  and  $V''$ , it follows  $x' = y'$  and  $x'' = y''$ .  $\blacktriangle$

**Definition B.23** Suppose  $V = V' \oplus V''$ . The projection  $p : V \rightarrow V'$  is the function  $p(x) = x'$ ,  $x \in V$ . The projection is a linear and continuous function.

**Example B.23** Let  $(x_1, \dots, x_n)$  be a basis for an  $n$ -dimensional vector space  $V$ . The sets  $V_i = \{\alpha_i x_i : \alpha_i = \text{scalars}\}$ ,  $i = 1, \dots, n$ , are linear subspaces of  $V$  whose only common point is  $0 \in V$ , so that  $V = \bigoplus_{i=1}^n V_i$ . The functions  $p_i : V \rightarrow V_i$  defined by  $p_i(x) = x_i$ ,  $x \in V$ ,  $i = 1, \dots, n$ , are projections of  $V$  on its linear subspaces  $V_i$ .  $\diamond$

### B.3.1 Completion of Metric Spaces

Convergent sequences in metric spaces are Cauchy sequences. However, the converse may or may not be true, as illustrated by Example B.22. If a Cauchy sequence in a metric space  $(M, d)$  is not convergent, it can be made convergent by adding points to  $M$  such that it becomes a complete space. Note that a normed linear space  $(V, \|\cdot\|)$  is a metric space  $(V, d)$  with the metric  $d(x, y) = \|x - y\|$ ,  $x, y \in V$ .

Following are two essential results on the completion of metric spaces that are given without proof, the completion theorem and the Weierstrass approximation theorem.

**Theorem B.19** (The completion theorem) *For any metric space  $(M, d)$  there exists a complete metric space  $(M', d')$  and a dense subset  $\tilde{M}$  of  $M'$  such that  $(M, d)$  and  $(\tilde{M}, d')$  are isometric. All completions of  $(M, d)$  are isometric, that is, they preserve the distance ([14], Theorem 2.1.6).*

**Theorem B.20** (The Weierstrass approximation theorem) *The set  $\mathcal{P}$  of polynomials on a closed and bounded interval  $[a, b]$  is dense in  $(C[a, b], \|\cdot\|_\infty)$ . Hence, any real-valued continuous function defined on  $[a, b]$  can be approximated to any degree of accuracy by a polynomial. Also, the normed linear space  $(C[a, b], \|\cdot\|_\infty)$  is a completion of the normed linear space  $(\mathcal{P}, \|\cdot\|_\infty)$  ([14], Theorems 2.2.1 and 2.2.3).*

**Example B.24** We have seen in Example B.22 that  $C[0, 1]$  with the norm  $\|x\|_1 = \|x\| = \int_0^1 |x(t)| dt$  is not complete. If we enlarge  $C[0, 1]$  to the space  $L^1[0, 1]$  of real-valued Lebesgue integrable functions defined on  $[0, 1]$ , then  $(L^1[0, 1], \|\cdot\|_1)$  is a Banach space. We will discuss in some details  $L^p$ -spaces,  $p \geq 1$ , in Sect. B.5.  $\diamond$

### B.3.2 Basis and Separability

**Definition B.24** Let  $(V, \|\cdot\|)$  be a normed linear space. A finite set  $\{x_1, \dots, x_n\}$  of elements in  $V$  is a basis for  $V$  if every  $x \in V$  admits a unique representation of the form  $x = \sum_{i=1}^n \alpha_i x_i$ . If  $\{x_1, \dots, x_n\}$  is a basis for  $V$ , we say that  $V$  has dimension  $n$ .

A countable set  $\{x_1, \dots, x_n, \dots\}$  of elements in  $V$  is a Schauder basis for  $V$  if every  $x \in V$  admits a unique representation of the form  $x = \sum_{i=1}^\infty \alpha_i x_i$ . If  $\{x_1, \dots, x_n, \dots\}$  is a Schauder basis for  $V$ , then  $\sum_{i=1}^\infty \alpha_i x_i = 0$  implies  $\alpha_i = 0$ ,  $\forall i$ .

**Definition B.25** A normed vector space  $V$  is said to be separable by a basis if it admits a finite or a Schauder basis.

Note that a topological space is separable if it has a countable dense subset and that a normed linear space that is separable by a basis is separable ([14], Theorem 3.3.6). For example, if  $V$  is  $n$ -dimensional, the set  $\{\sum_{i=1}^n \alpha_i x_i : \Re(\alpha_i), \Im(\alpha_i) \in \mathbb{Q}\}$  is countable and dense in  $V$ . We also note that separability by a basis and separability



do not coincide for Banach spaces ([14], Sect. 3.3.2), but they do for Hilbert spaces, as shown in a subsequent section.

*Example B.25* A Schauder basis for the space  $l^2$  of square summable sequences consists of the sequences  $(1, 0, 0, \dots)$ ,  $(0, 1, 0, \dots)$ ,  $\dots$  in  $\mathbb{R}$ .  $\diamond$

### B.3.3 Operators

Let  $(V, \|\cdot\|_V)$  and  $(W, \|\cdot\|_W)$  be normed linear spaces, and let  $T : V \rightarrow W$  be a function from  $V$  into  $W$ , referred to as an operator from  $V$  into  $W$ . If  $W$  is  $\mathbb{R}$  or  $\mathbb{C}$ , then  $T$  is said to be a functional.

**Definition B.26**  $T$  is a linear operator if  $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ ,  $x, y \in V$ , where  $\alpha, \beta \in \mathbb{F}$  are scalars. A linear operator is called bounded if there exists  $M > 0$  such that  $\|T(x)\|_W \leq M\|x\|_V$  for all  $x \in V$ . The norm of  $T$  is  $\|T\| = \inf\{M > 0 : \|T(x)\|_W \leq M\|x\|_V, x \in V\}$ .

*Example B.26* Let  $V = \mathbb{R}^q$ ,  $W = \mathbb{R}^{q'}$ , and  $T = \{t_{ij}\}$  a  $(q', q)$ -matrix with real-valued entries. Then  $T$  is a linear operator, and  $\|Tx\|^2 = \sum_{i=1}^{q'} \left(\sum_{j=1}^q t_{ij}x_j\right)^2 \leq \sum_{i=1}^{q'} \left(\sum_{j=1}^q t_{ij}^2\right)^{1/2} \left(\sum_{j=1}^q x_j^2\right)^{1/2} \leq q'q\bar{t}\|x\|$  by the Cauchy–Schwarz inequality, where  $\bar{t} = \max_{1 \leq i \leq q', 1 \leq j \leq q} |t_{ij}|$ .  $\diamond$

*Example B.27* Let  $V = W = C^1[0, 1]$  be the space of real-valued functions that are continuous and continuously differentiable with the norm  $\|x\|^2 = \int_0^1 x(t)^2 dt$  and let  $T$  be the differentiation operator. The operator  $T$  is unbounded since, for example,  $x(t) = \sin(n\pi t)$ ,  $t \in [0, 1]$ , is a member of  $C^1[0, 1]$ , has norm  $\|x\|^2 = 1/2$ , and its image  $Tx$  has norm  $\|Tx\|^2 = \int_0^1 n^2\pi^2 \cos^2(n\pi t) dt = n^2\pi^2/2$  so that  $\|Tx\|/\|x\| = (n\pi)^{1/2}$ .  $\diamond$

**Definition B.27** Let  $B(V, W)$  be the set of bounded linear operators  $T : V \rightarrow W$ . If  $W$  is  $\mathbb{R}$  or  $\mathbb{C}$ , then  $B(V, W)$  is the set of real/complex-valued bounded linear functionals, and is called the dual space of  $V$ .

**Theorem B.21** *The set  $B(V, W)$  with the pointwise operations*

$$\begin{aligned} (T_1 + T_2)(x) &= T_1(x) + T_2(x), \quad x \in V, \\ (\alpha T)(x) &= \alpha T(x), \quad x \in V, \end{aligned} \tag{B.13}$$

*is a linear space, where  $T, T_1, T_2 \in B(V, W)$  and  $\alpha$  is a scalar.*

*Proof* Since  $B(V, W)$  is closed under these operations, it is a linear space.  $\blacktriangle$

**Theorem B.22** *If  $T$  is continuous at a point  $x_0 \in V$ , it is continuous in  $V$ .*

*Proof* For  $x \neq x_0$  and  $z = y + x - x_0$ ,  $x, y \in V$ , we have  $T(z) = T(y) + T(x) - T(x_0)$  so that  $\|T(z) - T(x)\| = \|T(y) - T(x_0)\| \rightarrow 0$  if  $\|y - x_0\| \rightarrow 0$  since  $T$  is continuous at  $x_0$ . Hence,  $T$  is continuous at any  $x \in V$ .  $\blacktriangle$

**Theorem B.23** *A linear operator  $T$  is continuous if and only if it is bounded.*

*Proof* Suppose  $T$  is continuous but not bounded. Then, for each  $n \geq 1$ , there is  $x_n \in V$  such that  $\|T(x_n)\|_W > n\|x_n\|_V$  implying  $\|T(z_n)\|_W > 1$ , where  $z_n = x_n / (n\|x_n\|_V)$ . Since  $\lim_{n \rightarrow \infty} z_n = 0$ ,  $T$  is not continuous at the origin in contradiction with our assumption. If  $T$  is bounded, we have  $\|T(x) - T(y)\|_W \leq M\|x - y\|_V$ ,  $x, y \in V$ , implying that  $T$  is continuous.  $\blacktriangle$

**Theorem B.24** *If  $T$  is a bounded linear operator, then  $T \mapsto \|T\|$  is a norm on  $B(V, W)$ , where  $\|T\| = \sup_{\|x\|_V \leq 1} \|T(x)\|_W = \sup_{\|x\|_V = 1} \|T(x)\|_W$ .*

*Proof* That  $\|T\|$  is a norm on  $B(V, W)$  follows from (i)  $\|T\| \geq 0$  by definition, and  $\|T\| = 0$  if and only if  $\|T(x)\|_W = 0$  for all  $x \in V$  with  $\|x\|_V = 1$  implying  $T(x) = 0$  for  $\|x\|_V = 1$ , or  $T = 0$ , (ii)  $\|(\alpha T)(x)\|_W = |\alpha| \|T(x)\|_W$  by properties of  $T$  and of the norm in  $W$ , and (iii)  $\|(T_1 + T_2)(x)\|_W = \|T_1(x) + T_2(x)\|_W \leq \|T_1(x)\|_W + \|T_2(x)\|_W \leq \|T_1\| + \|T_2\|$  by properties of the norm in  $W$  and the definition of the norm on  $B(V, W)$ .  $\blacktriangle$

Note that, for  $x \neq 0$ , we have  $\|T\| \geq \|T(x/\|x\|_V)\|_W = \|T(x)\|_W / \|x\|_V$ , which implies  $\|T(x)\|_W \leq \|T\| \|x\|_V$ . The latter inequality is satisfied for  $x = 0$ , so that we have  $\|T(x)\|_W \leq \|T\| \|x\|_V$  for all  $x \in V$  ([15], Theorem II, p. 115).

**Theorem B.25** *If  $W$  is a Banach space, then  $B(V, W)$  with this norm is a Banach space.*

*Proof* We need to show that limits of Cauchy sequence  $\{T_n\}$  in  $B(V, W)$  are in this space. The idea of the proof is to construct the limit of  $\{T_n\}$  pointwise by noting that the sequence  $\{T_n(x)\} \subseteq W$  is Cauchy for any  $x \in V$  so that it has a limit  $T(x) \in W$  since  $W$  is Banach by assumption. To complete the proof, we need to show  $T \in B(V, W)$ . This part of the proof can be found in, for example, [15], (Theorem III, p. 115).  $\blacktriangle$

*Example B.28* Let  $B(V, W)$  be the set of  $(q', q)$ -matrices with real- or complex-valued entries, so that  $T = \{t_{ij}\}$ ,  $i = 1, \dots, q'$ ,  $j = 1, \dots, q$ , are operators in  $B(V, W)$  with the usual matrix operations,  $V = \mathbb{R}^q$  or  $\mathbb{C}^q$ , and  $W = \mathbb{R}^{q'}$  or  $\mathbb{C}^{q'}$ . Norms on the vector spaces  $V$  and  $W$  can be used to define norms for  $T$ , referred to as induced norms, and given by

$$\begin{aligned} \|T\| &= \max\{\|Tx\|_W : x \in V, \|x\|_V \leq 1\} = \{\|Tx\|_W : x \in V, \|x\|_V = 1\} \\ &= \{\|Tx\|_W / \|x\|_V : x \in V, \|x\|_V \neq 0\}. \end{aligned} \quad (\text{B.14})$$

This definition of  $\|T\|$  can be based on the  $p$ -norm for vectors (Example B.20). In the special case of Euclidean norm ( $p=2$ ), we have  $\|T\| = (\lambda_{\max}(TS))^{1/2}$ , where  $\lambda_{\max}(TS)$  denotes the largest eigenvalue of  $TS$  and  $S$  denotes the conjugate transpose of  $T$ . It is possible to define alternative matrix norms, for example, the maximum column or row sum of  $T$ , the norm  $\|T\| = (|t_{ij}|^p)^{1/p}$ , and many other norms ([17], Chap. 10).  $\diamond$

*Example B.29* Let  $T : L^2[a, b] \rightarrow L^2[a, b]$  be an (integral) operator defined by  $Tx(t) = \int_a^b k(t, s)x(s) ds$ , where the kernel  $k$  is a real-valued Lebesgue measurable

function on  $[a, b] \times [a, b]$  that is square integrable, that is,  $\int_{[a,b]^2} |k(t, s)|^2 dt ds < \infty$ . The operator is linear by properties of the Lebesgue integral and  $|Tx(t)| \leq \int_a^b |k(t, s)| x(s) ds \leq \left( \int_a^b k(t, s)^2 ds \right)^{1/2} \left( \int_a^b x(s)^2 ds \right)^{1/2}$  by properties of integrals and the Cauchy–Schwarz inequality. We have  $\|Tx\|^2 \leq \|x\| \int_{[a,b]^2} |k(t, s)|^2 dt ds$  by the definition of  $\|Tx\|$ , so that  $\|T\| \leq \int_{[a,b]^2} |k(t, s)|^2 dt ds$  by Theorem B.24.  $\diamond$

**Definition B.28** Let  $V$  and  $W$  be normed linear spaces and let  $T : V_0 \rightarrow W$  be a linear operator, where  $V_0$  is a subspace of  $V$ . We say that  $T$  is closed if, for every  $\{x_n\} \subseteq V_0$  converging to  $x \in V$ , the equality  $\lim_{n \rightarrow \infty} T(x_n) = y$  implies  $x \in V_0$  and  $T(x) = y$ .

## B.4 Hilbert Spaces

**Definition B.29** Let  $V$  be a linear space. A function  $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}, \mathbb{C}$  is called inner product if

$$\begin{aligned} \langle x, x \rangle &> 0, \quad \langle x, x \rangle = 0 \Leftrightarrow x = 0 && \text{(positive definite)} \\ \langle x, y \rangle &= \langle y, x \rangle^* && \text{(conjugate symmetry)} \\ \langle \alpha x, y \rangle &= \alpha \langle x, y \rangle, \quad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle && \text{(linearity)} \end{aligned} \tag{B.15}$$

for  $x, y, z \in V$  and  $\alpha \in \mathbb{R}$  or  $\mathbb{C}$ , where linearity in (B.15) refers to the first argument of the inner product. Two members  $x$  and  $y$  of  $V$  are said to be orthogonal if  $\langle x, y \rangle = 0$ , a property indicated by the notation  $x \perp y$ . A point  $x \in V$  is said to be orthogonal to a subset  $E$  of  $V$  if  $\langle x, y \rangle = 0$  for all  $y \in E$ , and we write  $x \perp E$ .

**Theorem B.26** The function  $x \mapsto \|x\|$  defined by

$$\|x\| = \sqrt{\langle x, x \rangle}, \quad x \in V, \tag{B.16}$$

is a norm on  $V$ . The norm is said to be induced by the inner product on  $V$ .

*Proof* The first two properties in (B.9) follow directly from the defining properties of the inner product. The triangle inequality results from the Cauchy–Schwarz inequality,

$$|\langle x, y \rangle| \leq \|x\| \|y\|, \quad x, y \in V, \tag{B.17}$$

which holds since  $\|\lambda x + y\|^2 = |\lambda|^2 \|x\|^2 + \lambda \langle x, y \rangle + \lambda^* \langle y, x \rangle + \|y\|^2 \geq 0$  for  $\lambda \in \mathbb{C}$  arbitrary and in particular for  $\lambda = -\langle y, x \rangle / \|x\|^2$ .  $\blacktriangle$

**Theorem B.27** The norm in (B.16) satisfies the identities

$$\begin{aligned} \|x + y\|^2 + \|x - y\|^2 &= 2\|x\|^2 + 2\|y\|^2 \quad (\text{parallelogram law}) \\ \|x\|^2 + \|y\|^2 &= \|x + y\|^2, \quad \text{if } \langle x, y \rangle = 0 \quad (\text{Pythagoras' theorem}). \end{aligned} \tag{B.18}$$

Also,  $f : V \rightarrow \mathbb{C}$  defined by  $f(x) = \langle x, y \rangle$  for an arbitrary but fixed  $y \in V$  is a linear and continuous function.

*Proof* These statements follow from the defining properties of the inner product and the Cauchy–Schwarz inequality. We also have  $|f(x) - f(x')| = |f(x - x')| = |\langle x - x', y \rangle| \leq \|x - x'\| \|y\|$  by the definition of  $f$  and the Cauchy–Schwarz inequality, which shows that  $f$  is continuous.  $\blacktriangle$

**Definition B.30** A linear space  $V$  endowed with an inner product is called a Hilbert space if it is a Banach space with respect to the norm induced by the inner product, that is, the norm in (B.16). We use the notation  $H$  for Hilbert spaces.

*Example B.30* Let  $L^2[a, b]$  be the vector space of all complex-valued Lebesgue measurable functions that are square integrable on  $[a, b]$ , endowed with the inner product  $\langle x, y \rangle = \int_a^b x(t)y(t)^* dt$ . The space  $L^2[a, b]$  with the norm induced by this inner product is a Hilbert space (Theorem B.66).  $\diamond$

*Example B.31* Let  $D$  be a bounded subset of  $\mathbb{R}^d$ . The set  $L^2(D)$  of real-valued Lebesgue measurable functions that are square integrable on  $D$  endowed with the inner product  $\langle x, y \rangle = \int_D x(t)y(t) dt$  is a Hilbert space (Example B.30). Related Hilbert spaces used to obtain weak solutions for partial differential equations are  $H^1(D) = \{x : D \rightarrow \mathbb{R} : x \in L^2(D), \partial x / \partial t_i \in L^2(D), i = 1, \dots, d\}$  endowed with the inner product  $\langle x, y \rangle_{H^1(D)} = \int_D (x(t)y(t) + \nabla x(t) \cdot \nabla y(t)) dt$ , where  $\nabla = (\partial / \partial t_1, \dots, \partial / \partial t_d)$ , and  $H_0^1(D) = \{x \in H^1(D) : x(t) = 0, t \in \partial D\}$  equipped with the same inner product as  $H^1(D)$ , where  $\partial D$  denotes the boundary of  $D$ . The norm induced on these spaces by the inner product is  $\|x\|_{H^1(D)}^2 = \int_D (x(t)^2 + |\nabla x(t)|^2) dt$ .  $\diamond$

**Theorem B.28** Let  $E$  be a closed and convex subset of a Hilbert space  $H$ . Then  $E$  contains an element with the smallest norm.

*Proof* Set  $\delta = \inf\{\|x\| : x \in E\}$  and let  $\{x_n\}$  be a sequence in  $E$  with the property  $\lim_{n \rightarrow \infty} \|x_n\| = \delta$ . Since  $E$  is a convex set,  $(x_m + x_n)/2 \in E$  and  $\|(x_m + x_n)/2\| \geq \delta$ . The parallelogram law gives  $\|x_m - x_n\|^2 = 2\|x_m\|^2 + 2\|x_n\|^2 - \|x_m + x_n\|^2 \leq 2\|x_m\|^2 + 2\|x_m\|^2 - 4\delta^2$  so that  $\|x_m - x_n\| \rightarrow 0$  as  $m, n \rightarrow \infty$  showing that  $\{x_n\}$  is Cauchy so that  $\lim_{n \rightarrow \infty} x_n = x \in H$ . Since  $E$  is closed, we have  $x \in E$ . From  $\lim_{n \rightarrow \infty} \|x_n\| = \delta$  and the continuity of the norm, it follows that  $\|x\| = \delta$ .  $\blacktriangle$

### B.4.1 Basis and Fourier Representations

We have seen that normed linear spaces can be decomposed in direct sums of linear subspaces. Similar decompositions are available for Hilbert spaces and can be used to construct Fourier representations for the members of these spaces.

**Definition B.31** Let  $V$  be a vector space endowed with an inner product. A collection  $\{e_1, e_2, \dots\} \subset V$  of orthogonal vectors with unit length, that is,  $\langle e_i, e_j \rangle = \delta_{ij}$ , is called

an orthonormal set. An orthonormal set is a basis in  $V$  if there is no other orthonormal set including it. The scalars  $c_i = \langle x, e_i \rangle$ ,  $x \in V$ , are called Fourier coefficients for  $x \in V$  with respect to the orthonormal sequence  $\{e_i\}$ .

**Theorem B.29** *Let  $W$  be a subspace of a Hilbert space  $H$ . For any  $x \in H$  we have  $x = x' + x''$ , where  $x' \in W$  and  $x'' \perp W$ . Moreover, the decomposition is unique.*

*Proof* If  $x \in W$ , the decomposition is valid with  $x'' = 0$ . If  $x \notin W$ , set  $E = \{z : z = x - x', x' \in W\}$ . Since  $W$  is a subspace of  $H$ ,  $E$  is convex and closed. Theorem B.28 states that there exists  $x'' = x - x' \in E$ ,  $x' \in W$ , with the smallest norm. Accordingly,  $\|x'' - \lambda y\|^2 \geq \|x''\|^2$  for all  $y \in W$  or, equivalently,  $-\lambda \langle y, x'' \rangle - \lambda^* \langle x'', y \rangle + |\lambda|^2 \|y\|^2 \geq 0$ , which gives  $|\langle y, x'' \rangle| \leq 0$  for  $\lambda = \langle x'', y \rangle / \|y\|^2$  implying  $\langle y, x'' \rangle = 0$ , that is,  $x'' \perp y$ . ▲

**Theorem B.30** *If  $\{v_i\}$  are mutually orthogonal vectors in a Hilbert space  $H$ , then*

$$\sum_{i=1}^{\infty} \|v_i\|^2 = \left\| \sum_{i=1}^{\infty} v_i \right\|^2 \quad (\text{Parseval's identity}) \quad (\text{B.19})$$

*provided the numerical series on the left side of the equality is convergent.*

*Proof* Let  $v_1, \dots, v_n$  be orthogonal vectors in  $H$ , that is,  $\langle v_i, v_j \rangle = 0$  for  $i \neq j$ . We have

$$\sum_{i=1}^n \|v_i\|^2 = \left\| \sum_{i=1}^n v_i \right\|^2 \quad (\text{Pythagoras' theorem}). \quad (\text{B.20})$$

Since  $H$  is a Hilbert space, the sequence of partial sums  $\{s_n = \sum_{i=1}^n v_i\}$  is Cauchy by the convergence of  $\sum_{i=1}^{\infty} \|v_i\|^2$ , so that the sequence  $\{s_n\}$  converges to  $\sum_{i=1}^{\infty} v_i$ . ▲

**Theorem B.31** *Let  $\{e_n\}$  be an orthonormal sequence in a Hilbert space  $H$ . The series  $\sum_{n=1}^{\infty} \alpha_n e_n$  is convergent in  $H$  if and only if  $\sum_{n=1}^{\infty} |\alpha_n|^2$  is convergent in  $\mathbb{R}$ . In case of convergence, we have  $\|\sum_{n=1}^{\infty} \alpha_n e_n\|^2 = \sum_{n=1}^{\infty} |\alpha_n|^2$ .*

*Proof* If  $\sum_{n=1}^{\infty} \alpha_n e_n$  is a convergent series, then the sequence  $s_n = \sum_{i=1}^n \alpha_i e_i$  of partial sums is Cauchy, that is,  $\|s_{n+k} - s_n\|^2 = \sum_{i=n+1}^{n+k} |\alpha_i|^2 \rightarrow 0$  as  $n \rightarrow \infty$  for  $k \geq 0$ . Hence,  $\sum_{i=1}^n |\alpha_i|^2$  is Cauchy and, therefore, convergent, so that  $\lim_{n \rightarrow \infty} \|s_n\|^2 = \lim_{n \rightarrow \infty} \sum_{i=1}^n |\alpha_i|^2 = \sum_{i=1}^{\infty} |\alpha_i|^2$ . ▲

**Example B.32** The functions  $\{e_n(t) = \exp(i2\pi nt), n = 0, \pm 1, \pm 2, \dots\}$ ,  $t \in [0, 1]$ , define an orthonormal set in the Hilbert space  $H = L^2[0, 1]$  since their inner product is  $\langle e_m, e_n \rangle = \int_0^1 e_m(t) e_n(t)^* dt = \int_0^1 \exp(i2\pi(m-n)t) dt = \delta_{mn}$ . Moreover, it is a basis in  $L^2[0, 1]$  ([16], Theorem 5.18.3). Accordingly,  $x \in L^2[0, 1]$  has the representation  $x = \sum_{n=-\infty}^{\infty} \alpha_n e_n$ , where  $\alpha_n = \int_0^1 x(t) e_n(t) dt$ . ◇

**Theorem B.32** *Let  $\{e_n\}$  be an orthonormal sequence in a Hilbert space  $H$ . For any  $x \in H$ , we have*

$$\left\| x - \sum_{i=1}^n \langle x, e_i \rangle e_i \right\|^2 = \|x\|^2 - \sum_{i=1}^n |\langle x, e_i \rangle|^2 \quad (\text{Bessel's equation}) \quad (\text{B.21})$$

so that

$$\sum_{i=1}^n |\langle x, e_i \rangle|^2 \leq \|x\|^2 \quad \text{and} \quad \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2 \quad (\text{Bessel's inequality}), \quad (\text{B.22})$$

which shows that  $\sum_{i=1}^{\infty} \langle e_i, x \rangle e_i$  is a convergent sequence.

*Proof* Since  $x - \xi$ ,  $\xi = \sum_{i=1}^n \langle x, e_i \rangle e_i$ , is orthogonal on the subspace spanned by  $(e_1, \dots, e_n)$ , we have  $\|x\|^2 = \|\xi\|^2 + \|x - \xi\|^2$  by Pythagoras' theorem, or  $\|x\|^2 = \sum_{i=1}^n |\langle x, e_i \rangle|^2 + \|x - \sum_{i=1}^n \langle x, e_i \rangle e_i\|^2$ , that is, the Bessel equation. The Bessel inequality follows from this equation since  $\|x - \sum_{i=1}^n \langle x, e_i \rangle e_i\|^2 \geq 0$ .  $\blacktriangle$

The Bessel equation and the Pythagoras theorem imply that  $\xi = \sum_{i=1}^n \langle x, e_i \rangle e_i$  is the best approximation of  $x \in H$  belonging to the subspace  $H_n$  spanned by  $(e_1, \dots, e_n)$ . For any vector  $z = \sum_{i=1}^n \alpha_i e_i$  in this subspace, we have  $\|x - z\|^2 = \|x - \xi + \xi - z\|^2 = \|x - \xi\|^2 + \|\xi - z\|^2 \geq \|x - \xi\|^2$ , that is, the error  $\|x - z\|$  for any  $z \in H_n$  is larger than  $\|x - \xi\| = \|x - \sum_{i=1}^n \langle x, e_i \rangle e_i\|$ .

**Theorem B.33** *Let  $W$  be a subspace of a Hilbert space  $H$  that is dense in  $H$ , and let  $\{e_n\}$  be an orthonormal basis for  $W$  that may or may not be finite. Then  $\{e_n\}$  is also an orthonormal basis for  $H$ .*

*Proof* If  $\{e_n\}$  is finite, the closure of  $W$  is  $H$  since finite dimensional spaces are complete. If  $\{e_n\}$  is not finite, we have the representation  $x = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$  for all  $x \in W$ . We need to show that any  $x \in H$  has a similar representation. Since the closure of  $W$  is  $H$ , for given  $\varepsilon > 0$  and an arbitrary  $x \in H$  there exists  $y \in W$  such that  $\|x - y\| = \|x - \sum_{n=1}^{\infty} \langle y, e_n \rangle e_n\| < \varepsilon$ . Since the series  $\sum_{n=1}^{\infty} \langle y, e_n \rangle e_n$  is convergent, there exists  $n_0$  such that  $\|y - y_n\| < \varepsilon$ ,  $n \geq n_0$ , where  $y_n = \sum_{i=1}^n \langle y, e_i \rangle e_i$ . Since  $x_n = \sum_{i=1}^n \langle x, e_i \rangle e_i$  is the best approximation of  $x$  in the subspace spanned by  $(e_1, \dots, e_n)$ , then  $\|x - x_n\| \leq \|x - y_n\|$ . We have  $\|x - x_n\| \leq \|x - y_n\| \leq \|x - y\| + \|y - y_n\| < 2\varepsilon$  for  $\varepsilon$  arbitrary so that  $x = \lim_{n \rightarrow \infty} \sum_{i=1}^n \langle x, e_i \rangle e_i = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ .  $\blacktriangle$

**Theorem B.34** *Every separable Hilbert space has an orthonormal basis.*

*Proof* If the space has finite dimension, the Gram–Schmidt algorithm can be used to construct such a basis. The construction is sequential. Let  $\{\xi_k\}$  be a finite or countable set of linearly independent vectors in a Hilbert space. The first member of the orthogonal basis is  $e_1 = \xi_1 / \|\xi_1\|$ . The second member of this basis is  $e_2 = (\xi_2 - \alpha e_1) / \|\xi_2 - \alpha e_1\|$ , where  $\alpha$  is the solution of  $\langle \xi_2 - \alpha e_1, e_1 \rangle = 0$ . The third member of the basis is the coordinate of  $\xi_3$  orthogonal to the subspace spanned by  $(\xi_1, \xi_2)$ , and so on. The proof of this property for Hilbert spaces that are not of finite dimension can be found in [14] (Theorem 3.4.9).  $\blacktriangle$

**Theorem B.35** Let  $\{e_i\}_{i \in I}$  be an arbitrary orthonormal set in a Hilbert space  $H$ , where  $I$  is an index set. The Fourier coefficients of an element  $x \in H$  with respect to  $\{e_i\}_{i \in I}$  are zero, except for countable many coefficients.

*Proof* Set  $c_i = \langle e_i, x \rangle$ ,  $i \in I$ , and note that  $\{i : c_i \neq 0\} = \cup_{p=1,2,\dots} \{i : |c_i| > 1/p\}$ . Consider a finite subset  $\{c_{i_1}, \dots, c_{i_n}\}$  of  $\{i : c_i \neq 0\}$ . The Bessel inequality gives  $\sum_{k=1}^n |c_{i_k}|^2 \leq \|x\|^2$  so that, if  $|c_{i_k}|^2 > 1/p$ ,  $k=1, \dots, n$ , then  $n \leq p\|x\|^2$  implying that the set  $\{i : |c_i|^2 > 1/p\}$  has less than  $p\|x\|^2$  members, so that  $\{i : c_i \neq 0\}$  is countable. This results gives meaning to the notation  $\sum_{i \in I} |c_i|^2$ , which is  $\sum_{i \in I} |c_i|^2 = \sum_{k=1}^{\infty} |c_{i_k}|^2$ , where  $\{c_{i_k}\}$ ,  $k=1, 2, \dots$ , are the non-zero Fourier coefficients.  $\blacktriangle$

**Theorem B.36** Let  $\{e_i\}_{i \in I}$  be an orthonormal set in a Hilbert space  $H$ . The Fourier coefficients of an element  $x \in H$  with respect to  $\{e_i\}_{i \in I}$  satisfy the inequality  $\sum_{i \in I} |c_i|^2 \leq \|x\|^2$ .

*Proof* For an arbitrary  $n$  we have  $\sum_{k=1}^n |c_{i_k}|^2 \leq \|x\|^2$  by Bessel's inequality. The above statement follows from this inequality by letting  $n \rightarrow \infty$ .  $\blacktriangle$

**Theorem B.37** Let  $E = \{e_i\}_{i \in I}$  be an orthonormal set in a Hilbert space  $H$ ,  $x \in H$ , and  $c_i = \langle x, e_i \rangle$  be Fourier coefficients. The series  $\sum_{i \in I} c_i e_i$  converges to an element  $x'$  such that  $x - x' \perp E$ .

*Proof* Let  $s_n = \sum_{k=1}^n c_{i_k} e_{i_k}$  denote the sequence of partial sums of  $\sum_{i \in I} c_i e_i$  corresponding to non-zero Fourier coefficients. This sequence is Cauchy since  $\|s_m - s_n\|^2 = \sum_{k=n+1}^m |c_{i_k}|^2$ ,  $m > n$ , and the series  $\sum_{k=1}^{\infty} |c_{i_k}|^2$  is convergent. Since  $H$  is a Hilbert space,  $\{s_n\}$  is also convergent. Let  $x' = \sum_{k=1}^{\infty} c_{i_k} e_{i_k}$  be the limit of  $\{s_n\}$ . We have  $\langle s_n, e_{i_p} \rangle = c_{i_p}$  provided  $p \leq n$ . The limit as  $n \rightarrow \infty$  gives  $\langle x', e_{i_p} \rangle = c_{i_p} = \langle x, e_{i_p} \rangle$  so that  $\langle x' - x, e_{i_p} \rangle = 0$  for all  $p \geq 1$  implying that  $x' - x$  is orthogonal to the subset of  $E$  spanned by those  $\{e_i\}$  with  $c_i \neq 0$ . For  $e_i \in E$  with  $c_i = 0$ , we have  $\langle x', e_i \rangle = 0 = \langle x, e_i \rangle$  or  $\langle x' - x, e_i \rangle = 0$  implying  $x' - x \perp E$ .  $\blacktriangle$

**Theorem B.38** An orthonormal set  $E = \{e_i\}_{i \in I}$  in a Hilbert space  $H$  is a basis for  $H$  if and only if  $x \perp E$  implies  $x = 0$ .

*Proof* Suppose  $E$  is a basis and  $x \perp E$  but  $x \neq 0$ , for example,  $\|x\| = 1$ . Hence,  $\{x\} \cup E$  is an orthonormal set contradicting the assumption that  $E$  is a basis. Conversely, suppose  $x \perp E$  implies  $x = 0$  but  $E$  is not a basis. Let  $F \supset E$  be a basis for  $H$ , so that there exists  $x \in F \setminus E$ ,  $x \neq 0$ . However,  $x \perp E$  so that  $x = 0$  which leads to contradiction.  $\blacktriangle$

**Theorem B.39** An orthonormal set  $E = \{e_i\}_{i \in I}$  in a Hilbert space  $H$  is a basis if and only if the linear space  $H_0$  spanned by it coincides with  $H$ .

*Proof* Suppose  $E$  is a basis for  $H$ . If  $H_0 \subset H$ , there exists a non-zero element in  $H \setminus H_0$  that is orthogonal on  $E$ , in which case  $E$  is not a basis, in contradiction to our assumption. Suppose now that  $H_0 = H$  and  $E$  is not a basis. Then, there exists an orthonormal basis  $F \supset E$ . Take  $y \in F \setminus E$  with  $\|y\| = 1$ . Since  $y \in H = H_0$ , it is an accumulation point for finite linear combinations of elements from  $E$ . Let  $\{z_n\}$

be such a sequence converging to  $y$ . We have  $\langle z_n, y \rangle = 0$  implying  $\langle y, y \rangle = \|y\|^2 = 0$  by taking the limit as  $n \rightarrow \infty$ , so that  $E$  is an orthonormal basis for  $H$ .  $\blacktriangle$

**Theorem B.40** *An orthonormal set  $E = \{e_i\}_{i \in I}$  in a Hilbert space  $H$  is a basis for  $H$  if and only if every  $x \in H$  admits the Fourier representation*

$$x = \sum_{i \in I} \langle x, e_i \rangle e_i, \quad x \in H. \quad (\text{B.23})$$

*Proof* If  $E$  is a basis, we have  $H_0 = H$  by the previous theorem, where  $H_0$  is the linear space spanned by  $E$ . We have shown that the series  $\sum_{i \in I} c_i e_i$ ,  $c_i = \langle x, e_i \rangle$ , is convergent with limit  $x'$  and that  $x - x' \perp E$ , so that  $x - x' \perp H_0 = H$ . Since  $x - x' \in H$ , we have  $\langle x - x', x - x' \rangle = 0$ , so that  $x' = x$ . Conversely, suppose the representation  $x = \sum_{i \in I} c_i e_i$  in (B.23) holds for all  $x \in H$  and  $E$  is not a basis. Then  $H_0 \subset H$  so that there exists  $x \in H$  with unit norm such that  $\langle x, e_i \rangle = 0$ ,  $i \in I$ . This observation and the representation of  $x$  imply  $x = 0$ . Hence,  $E$  must be a basis.  $\blacktriangle$

**Theorem B.41** *Let  $\{e_n\}$  be an orthonormal sequence in a Hilbert space  $H$ . The following statements are equivalent: (1)  $\{e_n\}$  is an orthonormal basis in  $H$ , (2) the equation  $\langle x, y \rangle = \sum_{n=1}^{\infty} \langle x, e_n \rangle \langle y, e_n \rangle$  holds for all  $x, y \in H$ , (3) the norm satisfies Parseval's identity in (B.19), that is,  $\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2$  holds for all  $x \in H$ , and (4) the sequence  $\{e_n\}$  is a total set, that is,  $\langle x, e_n \rangle = 0$  for all  $n \geq 1$  implies  $x = 0$  ([14], Theorem 3.4.10).*

*Example B.33* The set of orthonormal functions

$$(1/\sqrt{2\pi}, \cos(t)/\sqrt{\pi}, \sin(t)/\sqrt{\pi}, \dots, \cos(nt)/\sqrt{\pi}, \sin(nt)/\sqrt{\pi}, \dots) \quad (\text{B.24})$$

defines a basis for  $L^2[-\pi, \pi]$  so that we have the representation

$$x(t) = \alpha_0 + \sum_{n=1}^{\infty} (\alpha_n \cos(nt)/\sqrt{\pi} + \beta_n \sin(nt)/\sqrt{\pi}), \quad x \in L^2[-\pi, \pi], \quad (\text{B.25})$$

with coefficients  $\alpha_0 = \int_{-\pi}^{\pi} x(t) dt / \sqrt{2\pi}$ ,  $\alpha_n = \int_{-\pi}^{\pi} x(t) \cos(nt) dt / \sqrt{\pi}$ , and  $\beta_n = \int_{-\pi}^{\pi} x(t) \sin(nt) dt / \sqrt{\pi}$ ,  $n \geq 1$  by (B.23). Let  $x(t) = t$  and  $y(t) = 1$ ,  $t \in [-\pi, \pi]$ , be two members of  $L^2[-\pi, \pi]$ , so that they admit representations of the type in (B.25). The coefficients of these representations are  $\alpha_n = 0$ ,  $n \geq 0$ , and  $\beta_n = -2\sqrt{\pi}(-1)^n/n$ ,  $n \geq 1$ , for  $x(t)$  and  $\alpha_0 = \sqrt{2\pi}$ ,  $\alpha_n = \beta_n = 0$ ,  $n \geq 1$ , for  $y(t)$ , so that  $x(t) = -2 \sum_{n=1}^{\infty} (-1)^n \sin(nt)/n$  and  $y(t) = 1$ . The representations of  $x(t)$  and  $y(t)$  are consistent with Theorem B.41. For example,  $\langle x, y \rangle = \int_{-\pi}^{\pi} (t)(1) dt = 0$  by direct calculations and  $\sum_n \langle x, e_n \rangle \langle y, e_n \rangle = \langle x, 1/\sqrt{2\pi} \rangle \langle y, 1/\sqrt{2\pi} \rangle = 0$ . Also,  $\|x\|^2 = \int_{-\pi}^{\pi} t^2 dt = 2\pi^3/3$  and  $\sum_n |\langle x, e_n \rangle|^2 = \sum_{n=1}^{\infty} |\beta_n|^2 = 4\pi \sum_{n=1}^{\infty} (1/n^2) = (4\pi)(\pi^2/6) = 2\pi^3/3$ .  $\diamond$

**Theorem B.42** *The orthogonal complement  $W^\perp$  of a subset  $W \neq \emptyset$ , that is, the set  $W^\perp = \{y \in H : \langle y, x \rangle = 0 \text{ for all } x \in W\}$  is a closed linear subspace of  $H$  ([14],*



Theorem 3.5.5). If  $W_1$  and  $W_2$  are closed linear subspaces of a Hilbert space  $H$  such that  $W_1 \perp W_2$ , then (1) the representation  $x = y_1 + y_2$  in  $W_1 \oplus W_2$ ,  $y_k \in W_k$ , is unique and (2)  $W_1 \oplus W_2$  is a closed subspace of  $H$  ([14], Lemma 3.5.8). If  $W$  is a closed linear subspace of a Hilbert space  $H$ , then  $H = W \oplus W^\perp$ , a property referred to as the projection theorem ([14], Theorem 3.5.9).

## B.4.2 Linear Functionals

This section deals with functionals, that is, mappings from vector spaces to the fields underlying these spaces, which for our discussion are  $\mathbb{R}$  and  $\mathbb{C}$ . We also consider functionals defined on products of vector spaces.

**Definition B.32** A functional  $\varphi : V \rightarrow \mathbb{R}, \mathbb{C}$  is said to be linear if it satisfies the conditions in Definition B.26. The functional is bounded if there is a constant  $c > 0$  such that  $|\varphi(x)| \leq c\|x\|$  for all  $x \in V$ .

**Definition B.33** A mapping  $b : V \times V \rightarrow \mathbb{R}, \mathbb{C}$  is said to be a bilinear form if it is linear and anti-linear in the first and the second arguments, respectively, that is,

$$\begin{aligned} b(\alpha_1 x_1 + \alpha_2 x_2, y) &= \alpha_1 b(x_1, y) + \alpha_2 b(x_2, y) \\ b(x, \beta_1 y_1 + \beta_2 y_2) &= \beta_1^* b(x, y_1) + \beta_2^* b(x, y_2) \end{aligned} \quad (\text{B.26})$$

for  $x, x_1, x_2, y, y_1, y_2 \in V$  and arbitrary scalars  $\alpha_1, \alpha_2, \beta_1$ , and  $\beta_2$ .

**Definition B.34** A bilinear form is bounded if there is a constant  $c > 0$  such that

$$|b(x, y)| \leq c\|x\|\|y\|, \quad \forall x, y \in V. \quad (\text{B.27})$$

*Example B.34* The inner product is a bounded bilinear form. The conditions in (B.26) are satisfied by the defining properties of the inner product (B.15). For example,  $\langle x, \beta_1 y_1 + \beta_2 y_2 \rangle = \langle \beta_1 y_1 + \beta_2 y_2, x \rangle^* = (\beta_1 \langle y_1, x \rangle + \beta_2 \langle y_2, x \rangle)^* = \beta_1^* \langle x, y_1 \rangle + \beta_2^* \langle x, y_2 \rangle$  hold by conjugate symmetry and linearity in the first argument. The condition in (B.27) with  $c = 1$  follows from the Cauchy–Schwarz inequality.  $\diamond$

**Definition B.35** A bilinear form is positive definite, coercive, or elliptical if there exists a constant  $c' > 0$  such that

$$b(x, x) \geq c'\|x\|^2, \quad \forall x \in V. \quad (\text{B.28})$$

**Theorem B.43** (Riesz's representation theorem) If  $\varphi : H \rightarrow \mathbb{C}$  is a continuous linear functional and  $H$  is a Hilbert space, then there exists a unique vector  $z \in H$  such that  $\varphi(x) = \langle x, z \rangle$  for all  $x \in H$ .

*Proof* The geometric interpretation of the theorem is that  $\varphi(x)$  can be viewed as the projection  $\langle x, z \rangle$  of  $x$  on a unique element  $z \in H$  for all  $x \in H$ . Moreover,  $\|\varphi\| = \sup_{\|x\| \leq 1} |\varphi(x)| = \sup_{\|x\| \leq 1} |\langle x, z \rangle| \leq \|z\|$  by the Cauchy–Schwarz inequality.

We now prove Riesz's theorem. The kernel  $N = \{x \in H : \varphi(x) = 0\}$  of  $\varphi$  is a linear subspace in  $H$  by the linearity of  $\varphi$ . The subspace is closed since, if  $\{x_n\} \subseteq N$  converges to a limit  $x \in H$ , then  $\varphi(x) = \lim_{n \rightarrow \infty} \varphi(x_n) = 0$  by the continuity of  $\varphi$ , so that  $x \in N$ . If  $N = H$ , the proof is completed. Otherwise, we have  $H = N \oplus N^\perp$ . Take  $y \in N^\perp$  such that  $\|y\| > 0$  so that  $\varphi(y) \neq 0$ . For any  $x \in H$ , we have  $\varphi(x - y\varphi(x)/\varphi(y)) = 0$ . This implies  $\xi = x - y\varphi(x)/\varphi(y) \in N$ , so that  $x = y\varphi(x)/\varphi(y) + \xi$  and  $\langle x, y \rangle = \|y\|^2 \varphi(x)/\varphi(y)$  or  $\varphi(x) = \langle x, z \rangle$  for all  $x \in H$ , where  $z = y\varphi(y)^*/\|y\|^2$ . The vector  $z$  is unique. If  $z' \neq z$  has the same properties as  $z$ , we will have  $0 = \langle x, z \rangle - \langle x, z' \rangle = \langle x, z - z' \rangle$  implying  $\|z - z'\| = 0$  for  $x = z - z'$ , so that  $z = z'$ .  $\blacktriangle$

*Example B.35* A functional  $\varphi : L^2[a, b] \rightarrow \mathbb{C}$  is bounded and linear if and only if there exists  $y \in L^2[a, b]$  such that  $\varphi(x) = \int_a^b x(t)y(t)^* dt = \langle x, y \rangle$  for all  $x \in L^2[a, b]$ . The first part of this statement follows from the Riesz theorem since a linear operator is continuous if and only if it is bounded (Theorem B.23). The converse follows from the observations that  $\varphi$  is linear with norm  $\|\varphi\| = \sup_{\|x\| \leq 1} \left| \int_a^b x(t)y(t)^* dt \right|$  so that  $\|\varphi\| \leq \|x\| \|y\| \leq \|y\|$  by the Cauchy–Schwarz inequality. Hence, for  $x \neq 0$ , we have  $|\varphi(x/\|x\|)| \leq \|\varphi\| \leq \|y\|$ , which gives  $|\varphi(x)| \leq \|x\| \|y\|$ .  $\diamond$

**Theorem B.44** (Lax–Milgram theorem) *If  $b : H \times H \rightarrow \mathbb{R}, \mathbb{C}$  is a positive definite, bounded bilinear form and  $H$  is a Hilbert space, then for every bounded linear functional  $\varphi : H \rightarrow \mathbb{R}, \mathbb{C}$  there exists a unique  $y_\varphi$  such that  $\varphi(x) = b(x, y_\varphi)$  for all  $x \in H$  ([18], Sect. 3.1.3).*

*Proof* Riesz's representation theorem implies  $\varphi(x) = \langle x, z_\varphi \rangle$  for all  $x \in H$  and a unique  $z_\varphi \in H$  depending on  $\varphi$ . The Riesz theorem also implies that, for an arbitrary but fixed  $y \in H$ , we have  $b(x, y) = \langle x, z_y \rangle$  for all  $x \in H$  and a unique  $z_y \in H$  depending on  $y$  since  $b(\cdot, y)$  is a bounded linear functional. The coercivity of  $b$  is needed to show that the image  $T(H)$  of the operator  $T : H \rightarrow H$  defined by  $T(y) = z_y$  is dense in  $H$  so that any member of  $H$  is a limit of a sequence in  $T(H)$ . Let  $\{z_{\varphi,n}\} \subset T(H)$  be a sequence converging to  $z_\varphi \in H$  and set  $\{y_{\varphi,n} = T^{-1}(z_{\varphi,n})\} \subset H$ . The idea is to show that  $b(x, y_{\varphi,n})$  converges to  $b(x, y_\varphi = \lim_{n \rightarrow \infty} y_{\varphi,n})$  and that  $b(x, y_\varphi) = \langle x, z_\varphi \rangle$ , that is,  $b(x, y_\varphi) = \varphi(x)$  for all  $x \in H$ . For a complete proof see, for example, [18] (Sect. 3.1.3) or [19] (Theorem 13, p. 166).  $\blacktriangle$

### B.4.3 Weak Convergence

**Definition B.36** A sequence  $\{x_n\}$  in a Hilbert space  $H$  is weakly convergent with weak limit  $x$ , if for all  $y \in H$ , the sequence of real or complex numbers  $\langle x_n, y \rangle$  is convergent with limit  $\langle x, y \rangle$ , that is,  $\lim_{n \rightarrow \infty} \langle x_n, y \rangle = \langle x, y \rangle$  for all  $y \in H$ .

**Theorem B.45** *If  $x_n \rightarrow x$  weakly, then the weak limit  $x$  is unique.*

*Proof* Suppose  $\{x_n\}$  has two limits,  $x$  and  $x'$ , that is,  $\langle x_n, y \rangle \rightarrow \langle x, y \rangle$  and  $\langle x_n, y \rangle \rightarrow \langle x', y \rangle$  for all  $y \in H$  as  $n \rightarrow \infty$ . Since limits are unique in  $\mathbb{R}$  and  $\mathbb{C}$ , we have  $\langle x, y \rangle = \langle x', y \rangle$  or  $\langle x - x', y \rangle = 0$  for all  $y \in H$ , that is,  $x - x' \in H^\perp = \{0\}$ .  $\blacktriangle$

**Theorem B.46** *Let  $\{x_n\}$  be a sequence in a Hilbert space  $H$ . If  $x_n \rightarrow x$  strongly, that is,  $\|x_n - x\| \rightarrow 0$  as  $n \rightarrow \infty$ , then  $x_n \rightarrow x$  weakly.*

*Proof* We have  $|\langle x_n, y \rangle - \langle x, y \rangle| = |\langle x_n - x, y \rangle| \leq \|x_n - x\| \|y\|$  for all  $y \in H$  by properties of the inner product and the Cauchy–Schwarz inequality. This implies  $\langle x_n, y \rangle \rightarrow \langle x, y \rangle$  for all  $y \in V$  as  $n \rightarrow \infty$ .  $\blacktriangle$

**Theorem B.47** *Every weakly convergent sequence in a finite dimensional Hilbert space is also strongly convergent.*

*Proof* Let  $(e_1, \dots, e_k)$  be an orthonormal basis for an  $k$ -dimensional Hilbert space  $H$ , and let  $\{x_n\}$  be a sequence in  $H$ . Since  $x_n = \sum_{i=1}^k \langle x_n, e_i \rangle e_i$  and  $x = \sum_{i=1}^k \langle x, e_i \rangle e_i$ , we have  $\|x - x_n\|^2 = \sum_{i=1}^k |\langle x, e_i \rangle - \langle x_n, e_i \rangle|^2 \rightarrow 0$  as  $n \rightarrow \infty$  since  $x_n \rightarrow x$  weakly.  $\blacktriangle$

**Theorem B.48** *Every weakly convergent sequence in a Hilbert space is bounded ([14], Theorem 3.6.7).*

### B.4.4 Bounded Linear Operators

Let  $T$  be a member of  $B(H, H)$ , that is,  $T : H \rightarrow H$  is a bounded linear operator and  $H$  is a Hilbert space (Definition B.27). For simplicity, we denote  $B(H, H)$  by  $B(H)$ . The review in this section is based on [14] (Chap. 4).

**Definition B.37** If for  $T \in B(H)$  there exists a linear bounded operator  $\hat{T} \in B(H)$  such that  $\langle Tx, y \rangle = \langle x, \hat{T}y \rangle$  for all  $x, y \in H$ , then  $\hat{T}$  is called the adjoint operator of  $T$ . The existence of  $\hat{T}$  is guaranteed by the following theorem.

**Theorem B.49** *If  $T \in B(H)$ , then  $\hat{T}$  is a bounded linear operator.*

*Proof* The linearity of  $\hat{T}$  follows from its definition and properties of the inner product. The operator is bounded since  $\|\hat{T}x\|^2 = \langle \hat{T}x, \hat{T}x \rangle = \langle T(\hat{T}x), x \rangle \leq \|T(\hat{T}x)\| \|x\| \leq \|T\| \|\hat{T}x\| \|x\|$  by the definition of  $\hat{T}$ , the Cauchy–Schwarz inequality, and properties of the norm of  $T$ . We have  $\|\hat{T}x\| \leq \|T\| \|x\|$  for all  $x \in H$  by dividing with  $\|\hat{T}x\| \neq 0$  (the case  $\|\hat{T}x\| = 0$  is trivial). The latter inequality shows that  $\hat{T}$  is bounded and that  $\|\hat{T}\| \leq \|T\|$ .  $\blacktriangle$

**Definition B.38** An adjoint operator  $T \in B(H)$  is self-adjoint if  $T = \hat{T}$ .

**Theorem B.50** *The norms of the operators  $T$  and  $\hat{T}$  have the properties  $\|\hat{T}\| = \|T\|$  and  $\|\hat{T}T\| = \|T\|^2$ .*

*Proof* Arguments similar to those used to prove Theorem B.49 give  $\|T\| \leq \|\hat{T}\|$  so that  $\|T\| = \|\hat{T}\|$ . Also,  $\|Tx\|^2 = \langle Tx, Tx \rangle = \langle x, \hat{T}(Tx) \rangle \leq \|x\| \|\hat{T}T\| \|x\|$ ,

that is equal to  $\|\hat{T}T\|$  for  $\|x\|=1$  implying  $\|T\|^2 \leq \|\hat{T}T\|$ . Since  $\|\hat{T}T\| \leq \|\hat{T}\|\|T\| = \|T\|^2$ , we have  $\|\hat{T}T\| = \|T\|^2$ .  $\blacktriangle$

**Theorem B.51** *For any  $S, T \in B(H)$ , the adjoint of  $ST$  is  $\hat{T}\hat{S}$  and  $\hat{\hat{T}} = T$ , where  $\hat{T}$  denotes the adjoint of  $T$ .*

*Proof* These statements follow from the definition of the adjoint operator and properties of the inner product. For example,  $\langle Tx, y \rangle = \langle x, \hat{T}y \rangle = \langle \hat{T}y, x \rangle^* = \langle y, \hat{\hat{T}}x \rangle^* = \langle \hat{\hat{T}}x, y \rangle$ , so that  $\hat{\hat{T}} = T$  since  $x, y \in H$  are arbitrary.  $\blacktriangle$

**Theorem B.52** *Every operator  $T \in B(H)$  on a Hilbert space  $H$  induces the partition  $H = \ker(\hat{T}) \oplus \overline{T(H)}$ , where  $\ker(\hat{T}) = \{x \in H : \hat{T}x = 0\}$ .*

*Proof* We have  $H = \overline{T(H)} \oplus \overline{T(H)}^\perp$  by the projection theorem since the closure  $\overline{T(H)}$  of  $T(H)$  is a closed linear subspace. It remains to show that  $\overline{T(H)}^\perp = \ker(\hat{T})$ . For  $y \in \overline{T(H)}^\perp$ , we have  $\langle x, \hat{T}y \rangle = \langle Tx, y \rangle = 0$  for all  $x \in H$  implying  $\overline{T(H)}^\perp \subseteq \ker(\hat{T})$ . If  $y \in \ker(\hat{T})$ , we have  $\langle Tx, y \rangle = \langle x, \hat{T}y \rangle = \langle x, 0 \rangle = 0$  so that  $\ker(\hat{T}) \subseteq \overline{T(H)}^\perp$ .  $\blacktriangle$

**Theorem B.53** *If  $T \in B(H)$  is self-adjoint, then  $\langle Tx, x \rangle \in \mathbb{R}$ . The norm of  $T$  can be calculated from  $\|T\| = \sup\{\langle Tx, x \rangle : x \in H, \|x\| = 1\}$ .*

*Proof* The first statement holds since  $\langle Tx, x \rangle = \langle x, Tx \rangle = \langle Tx, x \rangle^*$ . A proof of the second part can be found in [14] (Theorem 4.1.11 and Assertion 4.1.12).  $\blacktriangle$

**Theorem B.54** *If  $T \in B(H)$  is self-adjoint and  $T(H)$  is dense in  $H$ , then  $T$  has an inverse operator defined on the image  $T(H)$  of  $T$ .*

*Proof* We have  $H = \ker(\hat{T}) \oplus T(H)^*$  by Theorem B.52 implying  $\ker(\hat{T}) = \{0\}$  since  $T(H)$  is dense in  $H$  by assumption. Since  $T$  is self-adjoint, we have  $\ker(T) = \ker(\hat{T}) = \{0\}$  so that  $T$  is injective, that is,  $x \neq x'$  implies  $Tx \neq Tx'$ , so that  $T^{-1} : T(H) \rightarrow H$  is well defined.  $\blacktriangle$

**Definition B.39** Let  $V$  and  $W$  be normed linear spaces. A bounded linear operator  $T : V \rightarrow W$ , that is, a member of  $B(V, W)$ , is said to be a compact or completely continuous operator if it maps any bounded subset  $A$  in  $V$  into a subset  $T(A)$  of  $W$  with compact closure  $\overline{T(A)}$  in  $W$ .

**Definition B.40** An operator  $T \in B(V, W)$  is said to be of finite rank if the dimension of  $T(V)$ , that is, the dimension of the linear subspace generated by the image of  $T$  in  $W$ , is finite.

**Theorem B.55** *A bounded linear operator of finite rank is a compact operator.*

*Proof* If  $A$  is a bounded subset of  $V$ , then  $T(A)$  is a bounded subset in  $W$ . Moreover,  $\overline{T(A)} \subset T(V)$  since  $T(V)$  is finite dimensional. Since  $\overline{T(A)}$  is bounded and closed in a finite dimensional space, it is compact.  $\blacktriangle$

**Theorem B.56** *Let  $H$  be an infinite dimensional, separable Hilbert space and*

$$Tx = \sum_{n=1}^{\infty} \alpha_n \langle x, e_n \rangle e_n, \quad x \in H, \quad (\text{B.29})$$

where  $\{e_n\}$  is an orthonormal basis for  $H$  and  $\{\alpha_n\}$  are complex numbers. If the set  $\{\alpha_n\}$  is bounded, the series in (B.29) is convergent and defines a linear operator  $T$  on  $H$ . Moreover,  $T$  exists and is bounded if and only if  $\{\alpha_n\}$  is bounded.

*Proof* For the first part of the proof, we use the following comparison test. If  $H$  is a Banach space and  $\{y_n\} \subseteq H$  is such that  $\|y_n\| \leq a_n$ ,  $n \geq n_0$ , and  $\sum_{n=1}^{\infty} a_n$  is convergent, then the sum  $\sum_{n=1}^{\infty} y_n$  is convergent. The test results from the inequality  $\|\sum_{k=n+1}^{n+p} y_k\| \leq \sum_{k=n+1}^{n+p} a_k$ ,  $n \geq n_0$ , that holds by the triangle inequality and the fact that  $\mathbb{R}$  and  $H$  are complete spaces.

The series  $\sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$  is convergent by Bessel's inequality for all  $x \in H$ . The comparison test implies that  $Tx = \sum_{n=1}^{\infty} \alpha_n \langle x, e_n \rangle e_n$  is convergent for all  $x \in H$  if  $\{\alpha_n\}$  is bounded. If the series is convergent, then  $T$  in (B.29) is defined and represents a linear operator on  $H$ .

We have  $\|Tx\|^2 = \sum_{n=1}^{\infty} |\alpha_n|^2 |\langle x, e_n \rangle|^2$ ,  $\|Te_n\| = |\alpha_n|$  for all  $n \in \mathbb{N}$ , and  $\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2$ , so that  $\|Tx\|^2 \leq \|x\|^2 \sum_{n=1}^{\infty} |\alpha_n|^2$ . Hence,  $T$  exists and is bounded if and only if the sequence  $\{\alpha_n\}$  is bounded. It can also be shown that  $T$  exists and is compact if and only if  $\lim_{n \rightarrow \infty} \alpha_n = 0$  ([14], Theorem 4.2.11).  $\blacktriangle$

*Example B.36* Let  $T : V \rightarrow V$  be a matrix operator with entries  $\{t_{ij} \in \mathbb{R}\}$ ,  $i, j = 1, \dots, n$ , defined by  $x \mapsto x' = Tx$ , where  $x'_i = \sum_{j=1}^n t_{ij} x_j$ ,  $i = 1, \dots, n$ , and  $V$  denotes the space of real-valued  $(n, n)$ -matrices. The operator is linear and bounded so that is a member of  $B(V)$ .  $\diamond$

*Example B.37* Let  $T : C[0, 1] \rightarrow C[0, 1]$  defined by  $Tx(t) = \int_0^1 k(t, s)x(s) ds$ , where the kernel  $k(\cdot, \cdot)$  is continuous in both arguments. Then  $T$  is linear and bounded, since  $k$  and  $x$  are continuous functions on bounded intervals and the integration domain is bounded.  $\diamond$

*Example B.38* Let  $T : C^2[0, \infty) \rightarrow C^2[0, \infty)$  be an operator with domain the space of continuous real-valued functions with continuous first and second derivatives defined by  $Tx(t) = x''(t)$ ,  $t \in [0, \infty)$ . The operator is linear but is not bounded since, for example,  $Tx(t) = 2$  for  $x(t) = t^2$  so that there is no  $M > 0$  such that  $\|Tx\| \leq M\|x\|$  for all  $x \in C^2[0, \infty)$ .  $\diamond$

*Example B.39* The operators in Examples B.36 and B.37 are adjoint. For  $T$  in Example B.36 we have  $\langle Tx, y \rangle = \sum_i (\sum_j t_{ij} x_j) y_i = \sum_j x_j (\sum_i t_{ij} y_i) = \langle x, \hat{T}y \rangle$ , where  $\hat{T}y = \sum_i t_{ij} y_i$ . Hence, the matrix representing  $\hat{T}$  is the transpose of the matrix representing  $T$ . For  $T$  in Example B.37 we have  $\langle Tx, y \rangle = \int_0^1 (\int_0^1 k(s, t)x(s) ds) y(t) dt$  that becomes  $\int_0^1 x(s) (\int_0^1 k(s, t)y(t) dt) ds = \langle x, \hat{T}y \rangle$  by changing the order of integration, where  $\hat{T}y(s) = \int_0^1 k(s, t)y(t) dt$ .  $\diamond$

### B.4.5 Spectral Theory

**Definition B.41** Let  $T : H \rightarrow H$  be a linear operator, where  $H$  is a Hilbert space. The resolvent set  $\rho(T)$  for  $T$  is a subset of complex numbers  $\{\zeta \in \mathbb{C}\}$  such that  $(\zeta I - T)^{-1}$  exists as a bounded linear operator on a subspace  $H_0$  of  $H$  dense in  $H$ . The spectrum  $\sigma(T)$  for  $T$  is the complement of  $\rho(T)$ , that is  $\sigma(T) = \mathbb{C} \setminus \rho(T)$ , and has three distinct components. The point spectrum  $\sigma_p(T)$  is the subset of  $\sigma(T)$  for which  $\zeta I - T$  cannot be inverted, so that it consists of the eigenvalues of  $T$ . The continuous spectrum  $\sigma_c(T)$  is the subset of  $\sigma(T)$  with the property  $(\zeta I - T)^{-1}$  exists as a densely defined and unbounded operator on  $H$ , and the residual spectrum  $\sigma_r(T)$  is the subset of  $\sigma(T)$  such that  $(\zeta I - T)^{-1}$  exists and is continuous, but its domain of definition is not dense in  $H$  ([14], 107 Chapter 5 and [15], Section IV.7).

*Example B.40* Let  $V$  denote the space of continuous functions  $x : [0, 1] \rightarrow \mathbb{R}$  with continuous first order derivative and piecewise continuous second order derivative such that  $x(0) = x(1) = 0$  and let  $T : V \rightarrow V$  be the linear operator  $Tx(t) = -d^2x(t)/dt^2$ . The general solution of  $Tx = \lambda x$  has the form  $x(t) = \alpha \sin(\sqrt{\lambda}t) + \beta \cos(\sqrt{\lambda}t)$ , where  $\alpha, \beta$  are arbitrary constants. The boundary condition  $x(0) = 0$  implies  $\beta = 0$  so that  $x(t) = \alpha \sin(\sqrt{\lambda}t)$ . The boundary condition  $x(1) = 0$  imposes the condition  $\alpha \sin(\sqrt{\lambda}) = 0$ . The option  $\alpha = 0$  is inadmissible since it corresponds to the trivial solution  $x(t) = 0$ ,  $t \in [0, 1]$ , so that we require  $\sin(\sqrt{\lambda}) = 0$ , which implies  $\lambda = \lambda_n = (n\pi)^2$ ,  $n = 1, 2, \dots$ , and  $x_n(t) = \sin(n\pi t)$ ,  $n = 1, 2, \dots$ ,  $t \in [0, 1]$ . The constants  $\{\lambda_n\}$  and the functions  $\{x_n(t)\}$  are the eigenvalues and the eigenfunctions of  $T$ . We also note that the eigenfunctions provide a basis for the set of square integrable functions defined on  $[0, 1]$  ([20], Sect. 2.2).  $\diamond$

*Example B.41* Let  $T = -d^2x(t)/dt^2$  be as in Example B.40, but assume that  $V$  is the vector space of continuous functions  $x : [0, \infty) \rightarrow \mathbb{R}$  with continuous first order derivative and piecewise continuous second order derivative that are square integrable such that  $x(0) = 0$ . The boundary condition  $x(0) = 0$  implies  $x(t) = \alpha \sin(\sqrt{\lambda}t)$  so that  $T$  has no eigenfunctions since  $\int_0^\infty x(t)^2 dt = \infty$ . To overcome this difficulty, we extend the classical definitions of eigenvalues and eigenfunctions.

We say that  $\lambda$  is a continuum eigenvalue for  $T$  if there exists a sequence of functions  $x_n(t)$  such that  $\lim_{n \rightarrow \infty} |(T - \lambda)x_n|/|x_n| = 0$ ;  $\lambda$  with this property belongs to the continuous spectrum of  $T$ . If the sequence  $x_n$  converges pointwise to a function  $x$ , then  $x$  is called a continuum eigenfunction of  $T$  ([20], Sect. 2.2). Since for  $\lambda > 0$  and  $t_n = (2n + 1/2)\pi\lambda^{-1/2}$  the sequence  $x_n(t) = \sin(\sqrt{\lambda}t)1(0 \leq t \leq t_n) + (1 - 2(t - t_n)^2)1(t_n \leq t \leq t_n + 1/2) + 2(t - t_n - 1)^21(t_n + 1/2 \leq t \leq t_n + 1)$  converges pointwise to  $x(t) = \sin(\sqrt{\lambda}t)$ ,  $t \geq 0$ , as  $n \rightarrow \infty$  and  $\lim_{n \rightarrow \infty} |(T - \lambda)x_n|/|x_n| = 0$ , the continuum eigenvalues and eigenfunctions are  $\lambda > 0$  and  $\sin(\sqrt{\lambda}t)$ ,  $t \geq 0$ .  $\diamond$

**Theorem B.57** If  $H$  is a Hilbert space and  $T \in B(H)$ , then the resolvent set  $\rho(T)$  is open in  $\mathbb{C}$  and the spectrum set  $\sigma(T)$  is closed in  $\mathbb{C}$ . Moreover,  $\sigma(T)$  is compact and included in the ball of radius  $\|T\|$  centered at 0 ([14], Theorems 5.1.7 and 5.1.8).

**Theorem B.58** *The eigenvalues of a self-adjoint linear operator  $T$  on a Hilbert space  $H$  are real-valued. Moreover, eigenvectors of distinct eigenvalues are orthogonal.*

*Proof* Let  $\lambda \in \mathbb{C}$  be an eigenvalue of  $T$  and  $x \neq 0$  an eigenvector corresponding to this eigenvalue, that is,  $Tx = \lambda x$ , so that  $\lambda \langle x, x \rangle = \langle Tx, x \rangle = \langle x, Tx \rangle = \lambda^* \langle x, x \rangle$  by properties of inner product and the definition of self-adjoint operators. Since  $\|x\| \neq 0$ , we have  $\lambda = \lambda^*$ .

Let  $\lambda, \mu$  be distinct eigenvalues of  $T$  and  $x, y$  eigenvectors corresponding to these eigenvalues, so that  $x, y \neq 0$ . We have  $\lambda \langle x, y \rangle = \langle Tx, y \rangle = \langle x, Ty \rangle = \mu \langle x, y \rangle$  implying  $(\lambda - \mu) \langle x, y \rangle = 0$ . Since  $\lambda \neq \mu$ , we have  $\langle x, y \rangle = 0$ .  $\blacktriangle$

*Example B.42* If  $T = \{t_{ij}\}$  is a real-valued symmetric matrix, then  $T$  is a self-adjoint operator and its eigenvalues are real-valued by Theorem B.58. This fact can be established by direct arguments.  $\diamond$

**Theorem B.59** *If  $T$  is a compact, self-adjoint linear operator defined on a separable Hilbert space  $H$ , then  $H$  admits an orthonormal basis  $\{e_n\}$  consisting of the eigenvectors of  $T$ , and  $T$  has the representation  $Tx = \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n$  for  $x = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$ , where  $\{\lambda_n\}$  denote the eigenvalues of  $T$  ([14], Theorem 5.2.8).*

**Theorem B.60** *If  $T$  is a compact linear operator on a Hilbert space  $H$  and  $\lambda \neq 0$  is an eigenvalue of  $T$ , the space  $H_\lambda = \{x \in H : Tx = \lambda x\}$  is a finite dimensional subspace of  $H$ .*

*Proof*  $H_\lambda$  is a linear space since it is closed under addition and scalar multiplication. Suppose  $H_\lambda$  is an infinite-dimensional space, and let  $\{x_n\}$  be an orthonormal sequence in  $H_\lambda$ . For  $m \neq n$ , we have  $\|Tx_m - Tx_n\|^2 = |\lambda|^2 \|x_m - x_n\|^2 = 2|\lambda|^2 > 0$ , where the latter equality holds by properties of  $\{x_n\}$ . This implies that  $\{Tx_n\}$  does not have any convergent subsequence, so that  $T$  cannot be compact.  $\blacktriangle$

We have seen that the series in (B.29) converges and defines a linear operator  $T$  if the sequence  $\{\alpha_n\}$  is bounded, the eigenvectors of  $T$  corresponding to distinct eigenvalues are orthogonal if  $T$  is self-adjoint, and the eigenvectors of a self-adjoint operator  $T$  corresponding to the same eigenvalue define a linear subspace.

*Example B.43* The linear space  $H_\lambda$  in Theorem B.60 associated with an eigenvalue  $\lambda$  of a linear compact operator  $T \in B(H)$  admits an orthogonal basis in  $H_\lambda$ .  $\diamond$

*Proof* Let  $x_1, x_2, \dots$  be the eigenvectors of  $T$  corresponding to an eigenvalue  $\lambda$  of this operator. Since  $Tx_1 = \lambda x_1, Tx_2 = \lambda x_2, \dots$ , we have  $T(\alpha_1 x_1 + \alpha_2 x_2 + \dots) = \lambda(\alpha_1 x_1 + \alpha_2 x_2 + \dots)$ , that is, any member of  $H_\lambda$  is an eigenvector of  $T$  corresponding to  $\lambda$  (see also [20], Sect. 2.2). The Gram–Schmidt algorithm can be used to map linearly  $x_1, x_2, \dots$  into a set of orthonormal vectors.  $\blacktriangle$

## B.5 $L^p$ Spaces

Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $p \geq 1$ , and  $\mathcal{B}$  the Borel  $\sigma$ -field on  $\mathbb{R}$ . Denote by  $\mathcal{L}^p$  the set of all real-valued measurable functions  $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$  that are  $p$ -integrable, that is,  $X^{-1}(\mathcal{B}) \subseteq \mathcal{F}$  and  $E[|X|^p] = \int |X|^p dP < \infty$ . We partition  $\mathcal{L}^p$  in classes of equivalence consisting of those members of  $\mathcal{L}^p$  that differ on events of probability 0, and denote by  $L^p(\Omega, \mathcal{F}, P)$  or  $L^p$  the collection of these classes of equivalence. We will not distinguish between the members of a class of equivalence. A similar construction holds for  $\mathbb{R}$ ,  $\mathbb{R}^d$ ,  $\mathbb{C}$ , and  $\mathbb{C}^d$ -valued measurable functions.

**Theorem B.61** *The space  $L^p(\Omega, \mathcal{F}, P)$  is a vector space over the field of real or complex numbers.*

*Proof* Let  $\alpha \in \mathbb{R}, \mathbb{C}$  and  $X, Y \in L^p$ . The functions  $\alpha X$  and  $X + Y$  are measurable since scalar multiplication and addition are continuous mappings. The functions  $|\alpha X|$  and  $X + Y$  are  $p$ -integrable since  $|\alpha X|^p = |\alpha|^p |X|^p$ , and  $|X + Y|^p \leq (|X| + |Y|)^p \leq 2^p \max\{|X|^p, |Y|^p\} \leq 2^p(|X|^p + |Y|^p)$ , so that  $|\alpha X|^p, |X + Y|^p \in L^1$ .  $\blacktriangle$

### B.5.1 Useful Inequalities

Let  $p \geq 1$  and define  $q \geq 1$  by the relationship  $1/p + 1/q = 1$ . A notable property of these numbers is the inequality

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}, \quad a, b \geq 0, \quad (\text{B.30})$$

that follows from  $\int_0^a \xi^{p-1} d\xi + \int_0^b \eta^{q-1} d\eta \geq ab$ , which results by comparing the areas under the graphs of function  $\eta(\xi) = \xi^{p-1}$ ,  $\xi \geq 0$ , and its inverse  $\xi(\eta) = \eta^{1/(p-1)} = \eta^{q-1}$ ,  $\eta \geq 0$ .

**Theorem B.62** *If  $X \in L^p$  and  $Y \in L^q$  with  $1/p + 1/q = 1$ , then  $XY \in L^1$ .*

*Proof* Since  $|XY| = |X||Y|$  and  $|X|, |Y| \geq 0$ , (B.30) with  $(|X|, |Y|)$  in place of  $(a, b)$  gives  $|XY| \leq |X|^p/p + |Y|^q/q \in L^1$ .  $\blacktriangle$

**Theorem B.63** *If  $X \in L^p$  and  $Y \in L^q$ , then*

$$\int |XY| dP \leq \left( \int |X|^p dP \right)^{1/p} \left( \int |Y|^q dP \right)^{1/q} \quad (\text{Hölder's inequality}). \quad (\text{B.31})$$

*Proof* Note first that  $E[|XY|] \leq (E[|X|^p])^{1/p} (E[|Y|^q])^{1/q}$  is an alternative form of (B.31). Also, Hölder's inequality for  $p = q = 2$  gives

$$E[|XY|] \leq \left( E[|X|^2] (E[|Y|^2]) \right)^{1/2} \quad (\text{Cauchy-Schwarz inequality}). \quad (\text{B.32})$$



The inequality (B.30) with  $a = |X|/(\int |X|^p dP)^{1/p}$  and  $b = |Y|/(\int |Y|^q dP)^{1/q}$  gives

$$\begin{aligned} |XY| &\leq (1/p)|X|^p \left( \int |X|^p dP \right)^{1/p-1} \left( \int |Y|^q dP \right)^{1/q} \\ &\quad + (1/q)|Y|^q \left( \int |Y|^q dP \right)^{1/q-1} \left( \int |X|^p dP \right)^{1/p}, \end{aligned}$$

so that  $\int XY dP \leq (1/p + 1/q) \left( \int |X|^p dP \right)^{1/p} \left( \int |Y|^q dP \right)^{1/q}$  by integration, that is, (B.31) since  $1/p + 1/q = 1$ .  $\blacktriangle$

**Theorem B.64** *If  $X, Y \in L^p$  and  $p > 1$ , then*

$$\left( \int |X + Y|^p dP \right)^{1/p} \leq \left( \int |X|^p dP \right)^{1/p} + \left( \int |Y|^p dP \right)^{1/p} \quad (\text{Minkowski's inequality}). \quad (\text{B.33})$$

*Proof* We have seen that  $X + Y \in L^p$  (Theorem B.61), so that the left side of (B.33) is meaningful. Note also that  $X \in L^p$  implies  $|X|^{p-1} \in L^q$  since  $(|X|^{p-1})^q = |X|^p \in L^1$ . This property and the Hölder inequality applied to the right side of the inequality  $|X + Y|^p = |X + Y|^{p-1}|X + Y| \leq |X + Y|^{p-1}|X| + |X + Y|^{p-1}|Y|$  gives (B.33). For example, since  $X + Y \in L^p$  we have  $|X + Y|^{p-1} \in L^q$  so that  $\int |X + Y|^{p-1}|X| dP \leq \left( \int |X|^p dP \right)^{1/p} \left( \int |X + Y|^p dP \right)^{1/q}$  by Hölder's inequality. Similar considerations for the term  $|X + Y|^{p-1}|Y|$  yield Minkowski's inequality.  $\blacktriangle$

## B.5.2 $L^p$ Spaces as Normed Spaces

We have seen that  $L^p$  is a linear space. It is shown that it is possible to define norms on  $L^p$  and an inner product on  $L^2$ .

**Theorem B.65** *The function  $X \mapsto \|X\|_p$ ,  $X \in L^p$ ,  $p \geq 1$ , defined by*

$$\|X\|_p = \left( \int |X|^p dP \right)^{1/p} = (E[|X|^p])^{1/p} \quad (\text{B.34})$$

*is a norm on  $L^p$ .*

*Proof* We have  $\|X\|_p \geq 0$  and  $\|X\|_p = 0$  if  $X = 0$   $P$ -a.s. and  $\|\alpha X\|_p = |\alpha| \|X\|_p$  by (B.34). The triangle inequality  $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$  follows from Minkowski's inequality. Hence,  $(L^p, \|\cdot\|_p)$  is a normed space.  $\blacktriangle$

**Theorem B.66** *The normed space  $(L^p, \|\cdot\|_p)$  is complete. Moreover, the collection of real/complex-valued, simple  $\mathcal{F}$ -measurable functions  $I_p$  is dense in  $L^p$ , so that any member of  $L^p$  can be represented by limits of sequences from  $I_p$  ([15], Sects. 10.2.1 and 10.2.2, [21], Theorem 5.1.1).*

**Definition B.42** Let  $X$  be a real-valued measurable function on  $(\Omega, \mathcal{F}, P)$ . The essential supremum of  $X$  is

$$\text{ess sup}(X) = \inf\{a \in \mathbb{R} : |X| \leq a \text{ a.s.}\}. \quad (\text{B.35})$$

If there exists  $M > 0$  such that  $|X| \leq M$  a.s., we say that  $X$  is essentially bounded.

**Theorem B.67** Let  $X$  be a real-valued measurable function on  $(\Omega, \mathcal{F}, P)$ . Then

$$\|X\|_\infty = \text{ess sup}(X) \quad (\text{B.36})$$

is a norm on the space  $L^\infty$  of real-valued measurable functions on  $(\Omega, \mathcal{F}, P)$  that are finite a.s.

*Proof* Since  $|X+Y| \leq |X|+|Y|$  and  $|\alpha X| = |\alpha||X|$ , we have  $\|X+Y\|_\infty \leq \|X\|_\infty + \|Y\|_\infty$  and  $\|\alpha X\|_\infty = |\alpha|\|X\|_\infty$  for any  $X, Y \in L^\infty$  and scalar  $\alpha$ . For example, (B.35) and (B.36) give  $|X| \leq \|X\|_\infty$  and  $|Y| \leq \|Y\|_\infty$  a.s., so that  $|X| + |Y| \leq \|X\|_\infty + \|Y\|_\infty$  a.s. implying that  $\|X\|_\infty + \|Y\|_\infty$  belongs to the set  $\{a \in \mathbb{R} : |X| + |Y| \leq a \text{ a.s.}\}$ . Hence,  $L^\infty$  is a vector space. Since  $\|X\|_\infty = 0$  holds if and only if  $X = 0$  a.s., we conclude that  $\|\cdot\|_\infty$  is a norm on  $L^\infty$ .  $\blacktriangle$

**Theorem B.68** If  $X$  is  $p$ -integrable, then

$$\lim_{p \rightarrow \infty} \left( \int |X|^p dP \right)^{1/p} = \text{ess sup}(X). \quad (\text{B.37})$$

*Proof* Let  $\text{ess sup}(X) = M$  and note that  $J_p = \left( \int |X|^p dP \right)^{1/p} \leq M$ . If  $M < \infty$ , we have  $J_p \geq \left( \int_{A_\varepsilon} |X|^p dP \right)^{1/p} > (M - \varepsilon)P(A_\varepsilon)^{1/p}$ , where  $A_\varepsilon = \{\omega \in \Omega : |X(\omega)| > M - \varepsilon\}$ . This implies  $J_p \geq M - \varepsilon$  for a sufficiently large  $p$ . Hence, for such  $p$  we have  $M - \varepsilon \leq J_p \leq M$  implying  $\lim_{p \rightarrow \infty} J_p = M$ . Suppose now  $M = \infty$ . Then  $|X|$  is not bounded on  $\Omega$ . Take  $\xi > 0$  such that  $A_\xi = \{\omega \in \Omega : |X(\omega)| > \xi\}$  has non-zero measure, that is,  $P(A_\xi) > 0$ . Then  $J_p \geq \left( \int_{A_\xi} |X|^p dP \right)^{1/p} > \xi P(A_\xi)^{1/p}$  so that  $J_p \geq \xi$  for a sufficiently large  $p$  implying  $J_p \geq \xi$  so that  $\lim_{p \rightarrow \infty} J_p = \infty = M$  since  $\xi$  is arbitrary ([15], Sect. 10.2).  $\blacktriangle$

**Theorem B.69** The function  $\langle \cdot, \cdot \rangle : L^2 \times L^2 \rightarrow \mathbb{R}$  defined by

$$\langle X, Y \rangle = \int XY dP = E[XY], \quad X, Y \in L^2 \quad (\text{B.38})$$

is an inner product on  $L^2$ .

*Proof* That  $\langle \cdot, \cdot \rangle$  satisfies the conditions in (B.26) follows from properties of expectation. The definition is meaningful since  $X, Y \in L^2$ , so that  $XY \in L^1$  by the Cauchy–Schwarz inequality or Theorem B.62. The norm  $X \mapsto \|X\| = (E[X^2])^{1/2}$  induced by this inner product coincides with that in (B.34) for  $p = 2$ . Similar considerations hold for complex-valued random variables. For  $\mathbb{C}$ -valued random variables, the inner product is  $\langle X, Y \rangle = E[XY^*]$  and  $\|X\| = E[|X|^2]^{1/2}$ .  $\blacktriangle$

## B.6 Orthogonal Polynomials

Let  $C^2(D) = \{f : D \rightarrow \mathbb{R}, D \subseteq \mathbb{R}^d\}$  be the set of real-valued continuous functions defined on a subset  $D$  of  $\mathbb{R}^d$  that have continuous first and second partial derivatives. We define on this linear space the inner products

$$\begin{aligned}\langle f, g \rangle &= \int_D f(x)g(x) dx \quad \text{and} \\ \langle f, g \rangle_w &= \int_D w(x)f(x)g(x) dx,\end{aligned}\tag{B.39}$$

where  $w : D \rightarrow \mathbb{R}$  is a positive weighting function.

For  $d = 1$  and  $D = [a, b]$ , consider the Sturm–Liouville differential equation

$$(p(x)y'(x))' - q(x)y(x) + \lambda w(x)y(x) = 0 \tag{B.40}$$

satisfying the boundary conditions

$$\begin{aligned}\alpha_1 y(a) + \alpha_2 y'(a) + \alpha_3 y(b) + \alpha_4 y'(b) &= 0 \\ \beta_1 y(a) + \beta_2 y'(a) + \beta_3 y(b) + \beta_4 y'(b) &= 0,\end{aligned}\tag{B.41}$$

where  $p', p, q, w \in C^0[a, b]$ ,  $p(x) > 0$ , and  $(\alpha_k, \beta_k)$ ,  $k = 1, \dots, 4$ , are constants. An alternative form of (B.40) is  $\mathcal{L}y = \lambda \mathcal{A}y$ , where  $\mathcal{L}y = -(py')' + qy$  and  $\mathcal{A}y = wy$  are operators defined on  $C^2[a, b]$ . That  $\mathcal{L}$  is self-adjoint results from the definition of the inner product. Since

$$\begin{aligned}\langle \mathcal{L}y, z \rangle &= - \int_a^b (p(x)y'(x))' z(x) dx + \int_a^b q(x)y(x)z(x) dx \\ &= - p(x)y'(x)z(x) \Big|_a^b + \int_a^b (p(x)y'(x)z'(x) + q(x)y(x)z(x)) dx \\ \langle y, \mathcal{L}z \rangle &= - p(x)z'(x)y(x) \Big|_a^b + \int_a^b (p(x)y'(x)z'(x) + q(x)y(x)z(x)) dx,\end{aligned}\tag{B.42}$$

we have  $\langle \mathcal{L}y, z \rangle = \langle y, \mathcal{L}z \rangle$  under the condition  $p(x)y'(x)z(x) \Big|_a^b = p(x)z'(x)y(x) \Big|_a^b$ , which holds if the boundary conditions in (B.41) are replaced with, for example,  $y(a) = y(b) = 0$ . In this case, the eigenvalues of  $\mathcal{L}y = \lambda \mathcal{A}y$  are real-valued and the eigenfunctions of this equation corresponding to distinct eigenvalues are orthogonal (Theorem B.58).

Consider (B.40) and (B.41) with  $p(x) = w(x)\beta(x)$ ,  $w'(x)/w(x) = \alpha(x)/\beta(x)$ ,  $w(x)\beta(x) \Big|_a^b = 0$ ,  $\alpha(x) = a_0 + a_1x$ ,  $\beta(x) = b_0 + b_1x + b_2x^2$ ,  $a_0, a_1, b_0, b_1, b_2$  constants,  $q(x) = 0$ , and  $y(a), y(b)$  finite. It can be shown that the eigenvalues and the eigenfunctions of the resulting Sturm–Liouville equation  $(w(x)\beta(x)y'(x))' + \lambda w(x)y(x) = 0$  are  $\lambda_n = -n((n+1)b_2 + a_1)$  and orthogonal polynomials  $y_n$  of degree  $n = 1, 2, \dots$ , for example,

Jacobi:  $[a, b] = [-1, 1]$ ,  $w(x) = (1-x)^r(1+x)^s$ ,  $\beta(x) = 1-x^2$ ,  
 where  $r$  and  $s$  are constants,  
 Legendre: Jacobi for  $r = s = 0$ ,  
 Chebyshev: Jacobi for  $r = s = 1/2$ ,  
 Laguerre:  $[a, b] \mapsto [0, \infty)$ ,  $w(x) = x^\zeta e^{-x}$ ,  $\zeta > -1$ , and  
 Hermite:  $[a, b] \mapsto (-\infty, \infty)$ ,  $w(x) = e^{-x^2}$ . (B.43)

We limit our discussion to Hermite polynomials and their use in the representation of random elements with finite variance. Additional information on orthogonal polynomials including the Askey scheme can be found elsewhere ([1], Sect. 4.1, [22], Chap. 1)

### B.6.1 Hermite Polynomials

The function  $\exp(-(x-a)^2/2)$  of  $a \in \mathbb{R}$  for a fixed  $x \in \mathbb{R}$  has the Taylor expansion

$$e^{-(a-x)^2/2} = \sum_{n=0}^{\infty} (-1)^n \frac{a^n}{n!} \frac{d^n(e^{-x^2/2})}{dx^n} \quad (\text{B.44})$$

about  $a = 0$ , so that

$$e^{ax-a^2/2} = \sum_{n=0}^{\infty} \frac{a^n}{n!} \left( (-1)^n e^{x^2/2} \frac{d^n(e^{-x^2/2})}{dx^n} \right), \quad a \in \mathbb{R}. \quad (\text{B.45})$$

**Definition B.43** The polynomial

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n(e^{-x^2/2})}{dx^n}, \quad x \in \mathbb{R}, \quad (\text{B.46})$$

is called the Hermite polynomial of degree  $n$ . For example, the Hermite polynomials of degrees  $n = 0, 1, 2$ , and  $3$  are  $H_0(x) = 1$ ,  $H_1(x) = x$ ,  $H_2(x) = x^2 - 1$ , and  $H_3(x) = x^3 - 3x$ . Note that  $e^{ax-a^2/2}$  in (B.45) can be written in the form

$$e^{ax-a^2/2} = \sum_{n=0}^{\infty} \frac{a^n}{n!} H_n(x). \quad (\text{B.47})$$

Let  $d\eta(x) = (2\pi)^{-1/2} \exp(-x^2/2) dx = d\Phi(x) = \phi(x) dx$ ,  $x \in \mathbb{R}$ , be the Gaussian measure with mean 0 and variance 1, and denote by  $L^2(\mathbb{R}, \mathcal{B}, d\eta(x))$  the Hilbert space of real-valued functions defined on the real line that are square integrable with respect to  $\eta$ , that is,  $\int_{\mathbb{R}} f(x)^2 d\eta(x) < \infty$  for  $f \in L^2(\mathbb{R}, \mathcal{B}, d\eta(x))$ . As previously,  $\mathcal{B}$  is a Borel  $\sigma$ -field on  $\mathbb{R}$ .

**Theorem B.70** The polynomials  $H_n(x)$ ,  $n \geq 0$ , are orthogonal in  $L^2(\mathbb{R}, \mathcal{B}, d\eta(x))$ . Moreover,  $E[H_m(X)H_n(Y)] = n!\rho^n\delta_{mn}$ , where  $(X, Y)$  are standard Gaussian variables with correlation coefficient  $\rho$ .

*Proof* The formula in B.47 with  $a = u$  and  $a = v$  gives

$$e^{(u+v)x - (u^2+v^2)/2} = \sum_{m,n=0}^{\infty} \frac{u^m v^n}{m!n!} H_m(x) H_n(x), \quad u, v \in \mathbb{R},$$

by multiplication. The expectation of the left side of this equation,  $\int_{-\infty}^{\infty} \exp((u+v)x - (u^2+v^2)/2) d\eta(x) = \exp(uv)$ , is a function of  $uv$  and so must be the expectation of its right side. The latter integration can be performed term by term by dominated convergence, so that we must have  $\int_{-\infty}^{\infty} H_m(x) H_n(x) d\eta(x) = 0$  for all  $m \neq n$ . This shows that the Hermite polynomials  $\{H_n(x)\}$  are orthogonal in  $L^2(\mathbb{R}, \mathcal{B}, d\eta(x))$ ,

$$e^{uv} = \sum_{n=0}^{\infty} \frac{(uv)^n}{(n!)^2} \int_{-\infty}^{\infty} H_n(x)^2 d\eta(x),$$

$\int_{-\infty}^{\infty} H_n(x)^2 d\eta(x) = n!$  since  $\exp(uv) = \sum_{n=0}^{\infty} (uv)^n / n!$ , and  $\{H_n(x)/\sqrt{n!}\}$  is an orthonormal sequence in  $L^2(\mathbb{R}, \mathcal{B}, d\eta(x))$ .

The Taylor expansion of  $E[H_m(X)H_n(Y)] = \int_{\mathbb{R}^2} H_m(x)H_n(y)\phi(x, y; \rho) dx dy$  about  $\rho = 0$  gives

$$E[H_m(X)H_n(Y)] = \sum_{k=0}^{\infty} \frac{\rho^k}{k!} \frac{d^k E[H_m(X)H_n(Y)]}{d\rho^k} \Big|_{\rho=0} = \frac{\rho^n}{n!} (n!)^2 \delta_{mn} \quad (\text{B.48})$$

where  $\phi(\cdot, \cdot; \rho)$  denotes the density of  $(X, Y)$ . The second equality holds since  $\partial\phi(x, y; \rho)/\partial\rho = \partial^2\phi(x, y; \rho)/(\partial x \partial y)$  and  $dH_m(x)/dx = mH_{m-1}(x)$  by (B.46), so that integration by parts gives

$$\frac{d}{d\rho} \int_{\mathbb{R}^2} H_m(x)H_n(y)\phi(x, y; \rho) dx dy = mn \int_{\mathbb{R}^2} H_{m-1}(x)H_{n-1}(y)\phi(x, y; \rho) dx dy.$$

The right side of this equation is 1 for  $m = n = 1$  and 0 otherwise at  $\rho = 0$ . Repeated differentiation of this equation gives  $d^n E[H_m(X)H_n(Y)]/d\rho^n \Big|_{\rho=0} = (n!)^2 \delta_{mn}$ .  $\blacktriangle$

**Theorem B.71** The collection of polynomials  $\{H_n(x)/\sqrt{n!}, n = 0, 1, \dots\}$  is an orthonormal basis in  $L^2(\mathbb{R}, \mathcal{B}, d\eta(x))$ , that is, every function  $g$  in this space has a unique series expansion

$$g(x) = \sum_{n=0}^{\infty} a_n \frac{H_n(x)}{\sqrt{n!}}, \quad \text{where } a_n = \frac{1}{\sqrt{n!}} \int_{-\infty}^{\infty} g(x) H_n(x) d\eta(x) \quad (\text{B.49})$$

and  $\|g\|^2 = \sum_{n=0}^{\infty} a_n^2$ .

*Proof* The essential steps of the proof can be found in [23] (Sect. 9.3). The expansion of  $g$  in (B.49) implies  $\|g\|^2 = \sum_{n=0}^{\infty} a_n^2$  (Theorem B.41).  $\blacktriangle$

Let  $X$  be a standard Gaussian variable and  $g : \mathbb{R} \rightarrow \mathbb{R}$  a measurable function such that  $E[g(X)^2] < \infty$ . Then  $Y = g(X) \in L^2(\mathbb{R}, \mathcal{B}, d\eta(x))$  has the unique representation

$$Y = g(X) = \sum_{n=0}^{\infty} a_n \frac{H_n(X)}{\sqrt{n!}}, \quad (\text{B.50})$$

where the coefficients  $a_n = E[g(X)H_n(X)]/\sqrt{n!}$  are given by (B.49). It is common in application to approximate  $Y = g(X)$  by truncations,

$$Y \simeq Y_r = \sum_{n=0}^r a_n \frac{H_n(X)}{\sqrt{n!}}, \quad (\text{B.51})$$

of the infinite series in (B.50). Note that  $E[Y_r^2]$  underestimates  $E[Y^2]$  by  $\sum_{n=r+1}^{\infty} a_n^2$  since  $E[g(X)^2] = \sum_{n=0}^{\infty} a_n^2 \geq \sum_{n=0}^r a_n^2 = E[Y_r^2]$  and that the sequence  $\{Y_r\}$  converges in m.s. to  $Y$ , that is,  $E[(Y_r - Y)^2] \rightarrow 0$  as  $r \rightarrow \infty$ .

*Example B.44* Consider a lognormal random variable defined by  $Y = g(X) = \exp(X)$ ,  $X \sim N(0, 1)$ . Since  $g$  is measurable and  $E[Y^2] < \infty$ ,  $Y$  admits a representation of the type in (B.50) with coefficients  $a_n = E[\exp(X)H_n(X)]/\sqrt{n!} = e^{1/2}/\sqrt{n!}$  so that  $Y_r = \sum_{n=0}^r e^{1/2}H_n(X)/n!$  provides an approximation for  $Y$ . We have  $Y_r \rightarrow Y$  in m.s. as  $r \rightarrow \infty$  by properties of Hermite polynomials. Direct calculations show the convergence  $\lim_{r \rightarrow \infty} E[Y_r^3] = E[Y^3]$ , so that  $E[Y^3]$  can be approximated by  $E[Y_r^3]$  for a sufficiently large  $r$  [2]. However,  $E[Y_r^k]$  may or may not provide satisfactory approximations for  $E[Y^k]$ ,  $k \geq 3$ , depending on the mapping  $X \mapsto Y = g(X)$ , as illustrated by the following example.  $\diamond$

*Example B.45* Let  $Y = g(X) = |X|$ , where  $X \sim N(0, 1)$ . Since  $g$  is measurable and  $E[Y^2] < \infty$ ,  $Y$  admits an expansion of the type in (B.50) with coefficients  $a_{2n} = (-1)^n / (\sqrt{2\pi} 2^{n-1} (2n-1)n!)$  and  $a_{2n+1} = 0$ , so that  $Y_{2r} = \sum_{n=0}^r a_{2n} H_{2n}(X)$  [2]. This approximation converges to  $Y$  in m.s. and, therefore, in probability, as  $r \rightarrow \infty$ .

However, higher order moments of  $Y_{2r}$  may not converge to the corresponding moments of  $Y$ . For example, the numerical sequence  $\alpha_{2r} = E[(Y_{2r} - Y_{2(r-1)})^4]$  diverges, that is,  $\lim_{r \rightarrow \infty} \alpha_{2r} = \infty$  ([24], Sect. 3.4, Example 2). Hence, the sequence  $\{Y_{2r}\}$  is not Cauchy in  $L^4(\mathbb{R}, \mathcal{B}, d\eta(x))$  so that  $\{Y_{2r}^4\}$  is not uniformly integrable ([25], Theorem 6.6.2). Accordingly, the sequence of moments  $\{E[Y_{2r}^4]\}$  does not converge to  $E[Y^4]$  ([26], Theorem 4.5.4). The practical implication of this result is that numerical approximations of  $E[Y^4]$  based on  $E[Y_r^4]$  are likely to be inaccurate irrespective of the order  $r$  of the approximation  $Y_{2r}$  of  $Y$ . Similarly, tails of the distribution of  $Y$  cannot be approximated by those of  $Y_{2r}$ .  $\diamond$

A slight generalization of the standard Hermite polynomials in (B.46) results by using the Gaussian measure  $d\eta^*(x) = (2\pi\gamma)^{-1/2} \exp(-x^2/(2\gamma))dx$ ,  $x \in \mathbb{R}$ , corresponding to a Gaussian variable with mean 0 and variance  $\gamma > 0$  ([23], Sect. 9.3), rather than a Gaussian variable with mean 0 and variance 1. For this measure, the Hermite polynomials are

$$H_n(x; \gamma) = (-\gamma)^n e^{x^2/(2\gamma)} \frac{d^n (e^{-x^2/(2\gamma)})}{dx^n}, \quad x \in \mathbb{R}, \quad (\text{B.52})$$

or  $H_n(x; \gamma) = H_n(u)\gamma^{-n/2}$  with  $u = x\gamma^{-1/2}$ . The statements of Theorems B.70 becomes

$$E[H_m(X; \gamma_x)H_n(Y; \gamma_y)] = n!\rho^n \gamma_x^{m/2} \gamma_y^{n/2} \delta_{mn}, \quad (\text{B.53})$$

where  $(X, Y)$  is a bivariate Gaussian vector with  $E[X] = E[Y] = 0$ ,  $E[X^2] = \gamma_x$ ,  $E[Y] = \gamma_y$ , and  $E[XY] = \rho\gamma_x^{1/2}\gamma_y^{1/2}$ . Theorem B.71 implies that  $\{H_n(x; \gamma)/\sqrt{\gamma^n n!}, n = 0, 1, \dots\}$  is an orthonormal system in  $L^2(\mathbb{R}, \mathcal{B}, d\eta^*(x))$  so that every function  $g$  in this space has the unique series expansion

$$g(x) = \sum_{n=0}^{\infty} a_n \frac{H_n(x; \gamma)}{\sqrt{n! \gamma^n}}, \quad \text{where} \quad a_n = \frac{1}{\sqrt{n! \gamma^n}} \int_{-\infty}^{\infty} g(y) H_n(y; \gamma) d\eta^*(y), \quad (\text{B.54})$$

and  $\|g\|^2 = \sum_{n=0}^{\infty} a_n^2$ .

The following identities,

$$\begin{aligned} x^n &= \sum_{k=0}^{[n/2]} \frac{n!}{(2k)!(n-2k)!} (2k-1)!! \gamma^k H_{n-2k}(x; \gamma) \\ H_{n+1}(x; \gamma) &= x H_n(x; \gamma) - \gamma n H_{n-1}(x; \gamma) \\ \frac{\partial H_n(x; \gamma)}{\partial x} &= n H_{n-1}(x; \gamma) \\ \frac{\partial H_n(x; \gamma)}{\partial \gamma} &= -\frac{1}{2} \frac{\partial^2 H_n(x; \gamma)}{\partial x^2}, \end{aligned} \quad (\text{B.55})$$

are for Hermite polynomials corresponding to the Gaussian measure  $d\eta^*(x) = (2\pi\gamma)^{-1/2} \exp(-x^2/(2\gamma))dx$ ,  $x \in \mathbb{R}$ , and are useful in calculations. The symbol  $[n/2]$  denotes the integer part of  $n/2$  and  $n!!$  is equal to  $n \cdot (n-2) \cdots 5 \cdot 3 \cdot 1$  for  $n > 0$  odd,  $n \cdot (n-2) \cdots 6 \cdot 4 \cdot 2$  for  $n > 0$  even, and 1 for  $n = -1, 0$ .

## B.6.2 Homogeneous Chaos

Let  $\{B(t), t \geq 0\}$  be a Brownian motion on a probability space  $(\Omega, \mathcal{F}, P)$  and let  $L^2[a, b]$  denote the Hilbert space of square integrable functions  $f: [a, b] \rightarrow$

$\mathbb{R}$ ,  $[a, b] \subseteq \mathbb{R}$ , with respect to the Lebesgue measure, that is,  $f \in L^2[a, b]$  implies  $\|f\| = \int_a^b f(t)^2 dt < \infty$ . The stochastic integral  $I(f) = \int_a^b f(t) dB(t)$  is a Gaussian random variable on  $(\Omega, \mathcal{F}, P)$  with mean  $E[I(f)] = 0$  and variance  $E[I(f)^2] = \int_a^b f(t)^2 dt = \|f\|^2$  (Sect. 4.4). The function  $\omega \mapsto \int_a^b f(t) dB(t, \omega)$ ,  $\omega \in \Omega$ , is  $\mathcal{F}$ -measurable, but is also measurable with respect to a smaller  $\sigma$ -field, the  $\sigma$ -field  $\mathcal{F}^B = \sigma(B(t) : a \leq t \leq b)$  generated by  $B$ . The review in this section follows developments in [23] (Sect. 9.4). Useful information on this topic can also be found in [27] (Sect. 2.2) and [28] (Chap. 6).

Denote by  $L^2(\Omega, \mathcal{F}^B, P) = L_B^2(\Omega)$  the Hilbert space of  $P$ -square integrable functions on  $\Omega$  that are  $\mathcal{F}^B$ -measurable. The stochastic integral  $I(f)$  is a member of  $L^2(\Omega, \mathcal{F}^B, P)$  and defines a mapping  $I : L^2[a, b] \rightarrow L^2(\Omega, \mathcal{F}^B, P)$  that is an isometry since the norms  $E[I(f)^2] = \int_a^b f(t)^2 dt$  and  $\|f\|^2 = \int_a^b f(t)^2 dt$  of  $I(f)$  and  $f \in L^2[a, b]$  are equal (Sect. 4.4).

Set  $J_0 = \mathbb{R}$ . Let  $\{J_n, n \geq 1\}$ , denote the closures in  $L^2(\Omega, \mathcal{F}^B, P)$  of the linear space spanned by real-valued constant functions and products  $I(f_1) \cdots I(f_k)$ ,  $f_1, \dots, f_k \in L^2[a, b]$ ,  $k = 1, \dots, n$ , respectively. By construction, we have  $J_0 \subset J_1 \subset \cdots \subset J_n \subset \cdots \subset L^2(\Omega, \mathcal{F}^B, P)$ . Note that  $J_n$ ,  $n \geq 1$ , are infinitely dimensional spaces.

**Theorem B.72** *The union  $\cup_{n=1}^{\infty} J_n$  is dense in  $L^2(\Omega, \mathcal{F}^B, P)$ .*

**Proof** See Theorem 9.4.5 in [23]. The theorem implies that an arbitrary member of  $L^2(\Omega, \mathcal{F}^B, P)$  can be approximated by a sequence in  $\cup_{n=1}^{\infty} J_n$  to the desired accuracy. Also, the sequence  $\{J_n, n \geq 0\}$  of subspaces in  $L^2(\Omega, \mathcal{F}^B, P)$  can be replaced with  $K_0 = J_0 = \mathbb{R}$  and  $\{K_n, n \geq 1\}$ , where  $K_n$  denotes the orthogonal complement of  $J_{n-1}$  in  $J_n$  so that  $J_n = J_{n-1} \oplus K_n$  can be represented by the orthogonal direct sum of  $J_{n-1}$  and  $K_n$ . ▲

**Definition B.44** The members of  $K_n$  are called homogeneous chaoses of order  $n$ .

**Theorem B.73** *The representation  $L^2(\Omega, \mathcal{F}^B, P) = \oplus_{n=0}^{\infty} K_n$  holds, so that each member  $\phi \in L^2(\Omega, \mathcal{F}^B, P)$  has a unique homogeneous chaos expansion*

$$\phi = \sum_{n=0}^{\infty} \phi_n, \quad \phi_n \in K_n, \quad (\text{B.56})$$

and  $\|\phi\|^2 = \sum_{n=0}^{\infty} \|\phi_n\|^2$ , where  $\|\cdot\|$  is the norm in  $L_B^2(\Omega)$ .

*Proof* See Theorem 9.4.7 in [23]. We illustrate the expansion in (B.56) by two elementary examples.

If  $\phi = \alpha + \beta I(f)$ ,  $\alpha, \beta \in \mathbb{R}$ ,  $f \in L^2[a, b]$ , then  $\phi \in J_1$ ,  $\phi_0 = \alpha \in K_0$ , and  $\phi_1 = \beta I(f) \in K_1$ . Note that  $E[(\phi - \alpha)c] = E[\phi_1 c] = c\beta E[I(f)] = 0$  for  $c \in K_0 = \mathbb{R}$  arbitrary, so that  $\phi_1$  is orthogonal on  $K_0$ .

If  $\phi = I(f)^2 - \|f\|^2$ , then  $\phi \in J_2$ ,  $E[\phi c] = 0$  for  $c \in K_0 = \mathbb{R}$  arbitrary implying  $\phi \notin K_0$ , and  $E[\phi I(g)] = E[I(f)^2 I(g)] - \|f\|^2 E[I(g)] = 0$ ,  $g \in L^2[a, b]$ , implying  $\phi \notin K_1$ . The latter equality holds since  $E[I(g)] = 0$  and  $E[I(f)^2 I(g)] = \int_{[a,b]^3} f(s)$



$f(t)g(u) dB(s) dB(t) dB(u) = 0$  by properties of multidimensional Wiener–Itô integrals discussed in the following section. Since  $\phi$  is a square of a stochastic integral, it must be in  $K_0 \cup K_1 \cup K_2$ , so that  $\phi = \sum_{n=0}^2 \phi_n$ . Since  $\phi$  has no projections on  $K_0$  and  $K_1$ , then  $\phi_0 = \phi_1 = 0$  and  $\phi = \phi_2 = I(f)^2 - \|f\|^2$ .  $\blacktriangle$

*Example B.46* Let  $I(f) = \int_a^b f(t) dB(t)$  be an Itô integral, where  $f \in L^2[a, b]$  and  $B$  denotes a Brownian motion. Then,  $I(f)^2 - \|f\|^2 \in K_2$  is a polynomial chaos of order  $n = 2$  that coincides with the double integral  $\int_a^b \int_a^b f(s)f(t) dB(s) dB(t)$ , referred to as the Wiener–Itô integral (Sect. B.6.3).  $\diamond$

*Proof* Set  $X(t) = \int_a^t f(s) dB(s)$ , or equivalently,  $dX(t) = f(t) dB(t)$  with  $X(a) = 0$ ,  $t \in [a, b]$ . The integral form of Itô’s formula in (Sect. 5.2) applied to  $X(t)^2$  gives  $X(b)^2 = 2 \int_a^b X(t) dX(t) + \int_a^b f(t)^2 dt$  so that  $2 \int_a^b f(t) X(t) dB(t) = I(f)^2 - \|f\|^2$ . Theorem B.78, stated in a subsequent section, implies

$$\int_a^b \int_a^b f(s)f(t) dB(s) dB(t) = 2 \int_a^b f(t) \left[ \int_a^t f(s) dB(s) \right] dB(t)$$

or

$$\int_a^b \int_a^b f(s)f(t) dB(s) dB(t) = 2 \int_a^b f(t) X(t) dB(t),$$

so that we have  $I(f)^2 - \|f\|^2 = \int_a^b \int_a^b f(s)f(t) dB(s) dB(t)$ .  $\blacktriangle$

*Example B.47* Let  $I(f) = \int_a^b f(t) dB(t)$  as in the previous example. Then

$$I(f)^n = H_n(I(f); \|f\|^2) + \frac{n(n-1)}{2} H_{n-2}(I(f); \|f\|^2) + \cdots \quad (\text{B.57})$$

by the first identity in (B.55) with  $I(f)$  and  $\|f\|^2$  in place of  $x$  and  $\gamma$ , respectively.  $\diamond$

**Definition B.45** Let  $\{e_k, k = 1, 2, \dots\}$  be an orthonormal basis in  $L^2[a, b]$  and consider a sequence  $\{n_k, k = 1, 2, \dots\}$  of nonnegative integers with finite sum. Set

$$\mathcal{H}_{n_1, n_2, \dots} = \prod_{k=1}^{\infty} \frac{1}{\sqrt{n_k!}} H_{n_k}(\tilde{e}_k) \quad (\text{B.58})$$

where  $\tilde{e}_k = I(e_k) = \int_a^b e_k(t) dB(t)$ . Since  $n_k \geq 0$  and  $\sum_{k=1}^{\infty} n_k < \infty$ , the product in the defining formula for  $\mathcal{H}_{n_1, n_2, \dots}$  has a finite number of terms.

**Theorem B.74** The collection of functions  $\{\mathcal{H}_{n_1, n_2, \dots}, n_1 + n_2 + \cdots = n\}$  is an orthonormal basis for  $K_n$ .

*Proof* See Lemma 9.5.3 and Theorem 9.5.4 in [23]. The theorem states that any member of  $K_n$ , that is, any polynomial chaos of order  $n$ , can be expressed as a linear form of  $\mathcal{H}_{n_1, n_2, \dots}$  with indices  $n_1, n_2, \dots$  such that  $n_1 + n_2 + \cdots = n$ .

For example, the orthonormal basis for  $K_0$ ,  $K_1$ ,  $K_2$ , and  $K_3$  are a real constant,  $\{H_1(\tilde{e}_k)\}$  for  $k = 1, 2, \dots$ ,  $\{H_1(\tilde{e}_k)H_1(\tilde{e}_l)\}$  and  $\{H_2(\tilde{e}_k)\sqrt{2!}\}$  for  $k, l = 1, 2, \dots, k \neq$

$l$ , and  $\{H_1(\tilde{e}_k)H_1(\tilde{e}_l)H_1(\tilde{e}_q)\}$ ,  $\{H_2(\tilde{e}_k)H_1(\tilde{e}_l)/\sqrt{2!}\}$ , and  $\{H_3(\tilde{e}_k)/\sqrt{3!}\}$  for  $k, l, q = 1, 2, \dots$ ,  $k \neq l \neq q$ , respectively.  $\blacktriangle$

**Example B.48** Let  $f \in L^2[a, b]$  so that it admits the representation  $f = \sum_{k=1}^{\infty} a_k e_k$  with  $a_k = \langle f, e_k \rangle = \int_a^b f(x) e_k(x) dx$ . Since  $I(f) \in K_1$ , it admits the representation  $I(f) = \sum_{k=1}^{\infty} a_k I(e_k) = \sum_{k=1}^{\infty} a_k \tilde{e}_k = \sum_{k=1}^{\infty} a_k H_1(\tilde{e}_k)$  by Theorem B.74, where  $\tilde{e}_k = I(e_k)$ . The representation is consistent with the fact that  $\{H_1(\tilde{e}_k) = \tilde{e}_k\}$  is an orthonormal basis for  $K_1$  (Theorem B.74).  $\diamond$

**Theorem B.75** The functions  $\{\mathcal{H}_{n_1, n_2, \dots}, n_1 + n_2 + \dots = n, n = 0, 1, \dots\}$  define an orthonormal basis for  $L^2(\Omega, \mathcal{F}^B, P)$  so that every  $\phi \in L^2(\Omega, \mathcal{F}^B, P)$  admits the representation

$$\phi = \sum_{n=0}^{\infty} \sum_{n_1+n_2+\dots=n} a_{n_1, n_2, \dots} \mathcal{H}_{n_1, n_2, \dots}, \quad (\text{B.59})$$

where  $n_1, n_2, \dots \geq 0$  and  $a_{n_1, n_2, \dots} = E[\phi \mathcal{H}_{n_1, n_2, \dots}]$ .

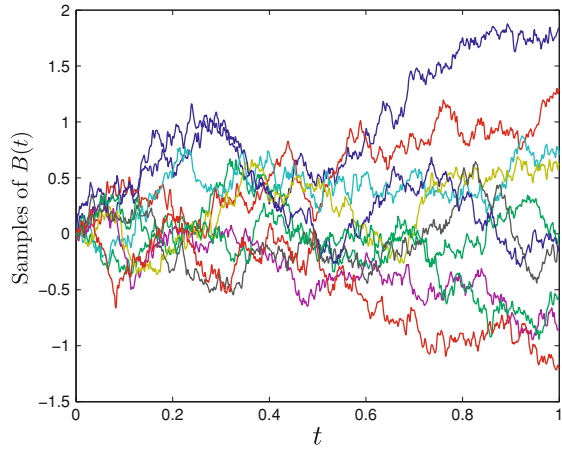
*Proof* See Theorem 9.5.7 in [23]. Note that  $L^2(\Omega, \mathcal{F}^B, P)$  is much larger than the space defined by measurable mappings of a Gaussian variable as considered in (B.50), and that the representation in (B.59) is a summation of Hermite polynomials  $\{H_{n_k}(\tilde{e}_k)\}$ ,  $\{H_{n_k}(\tilde{e}_k)H_{n_l}(\tilde{e}_l), \dots\}$  such that  $n_k \geq 0$ ,  $n_1 + n_2 + \dots = n$ , and  $n \geq 0$ , that is,

$$\begin{aligned} \phi &= a_0 + \sum_{k=1}^{\infty} a_{1,k} H_1(\tilde{e}_k) + \sum_{k,l=1, k \neq l}^{\infty} a_{2,kl} H_1(\tilde{e}_k) H_1(\tilde{e}_l) \\ &\quad + \sum_{k=1}^{\infty} a_{2,kk} H_2(\tilde{e}_k)/\sqrt{2!} + \dots \end{aligned} \quad (\text{B.60})$$

where  $\{\tilde{e}_k\}$  are independent Gaussian variables. In the expression of  $\phi$ , the first term  $a_0$  is a member of  $\mathcal{H}_{n_1, n_2, \dots}$  with  $n = 0$ , the second term  $\sum_{k=1}^{\infty} a_{1,k} H_1(\tilde{e}_k)$  includes members of  $\mathcal{H}_{n_1, n_2, \dots}$  with a single non-zero  $n_k = 1, 2, \dots$ , and the latter two terms  $\sum_{k,l=1, k \neq l}^{\infty} a_{2,kl} H_1(\tilde{e}_k) H_1(\tilde{e}_l)$  and  $\sum_{k=1}^{\infty} a_{2,kk} H_2(\tilde{e}_k)/\sqrt{2!}$  consist of members of  $\mathcal{H}_{n_1, n_2, \dots}$  with  $n = 2$ .  $\blacktriangle$

**Example B.49** Let  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a measurable function, and  $X = h(G_1, G_2)$ , where  $G_k$ ,  $k = 1, 2$ , are independent  $N(0, 1)$ . The random variables  $G_k$  can be viewed as two stochastic integrals  $I(f_k) = \int_a^b f_k(t) dB(t)$  corresponding to two functions  $f_k \in L^2[a, b]$  such that  $\int_a^b f_k(t)^2 dt = 1$ ,  $k = 1, 2$ , and  $\int_a^b f_1(t) f_2(t) dt = 0$  since  $E[I(f_k)] = 0$  and  $E[I(f_1)I(f_2)] = \int_a^b f_1(t) f_2(t) dt$  (Sect. 4.4). The random variable  $X$  admits the approximate representation

**Fig.B.1** Samples of  $B(t)$  using the first 600 terms of the series in (B.62)



$$\begin{aligned}
 X \simeq & a_0 + a_1 H_1(G_1) + a_2 H_1(G_2) + a_{11} \frac{H_2(G_1)}{\sqrt{2!}} \\
 & + 2a_{12} H_1(G_1) H_1(G_2) + a_{22} \frac{H_2(G_2)}{\sqrt{2!}}
 \end{aligned} \quad (\text{B.61})$$

by retaining from (B.59) terms of order  $n \leq 2$ , where  $a_0 = E[h(G_1, G_2)]$ ,  $a_k = E[h(G_1, G_2)H_1(G_k)]$ ,  $k = 1, 2$ ,  $a_{kk} = E[h(G_1, G_2)H_2(G_k)]/\sqrt{2!}$ ,  $k = 1, 2$ , and  $a_{12} = E[h(G_1, G_2)H_1(G_1)H_1(G_2)]$ .  $\diamond$

*Example B.50* Let  $\phi(t) = B(t)$ ,  $0 \leq t \leq 1$ , so that the coefficients  $a_{n_1, n_2, \dots}$  of the representation in (B.59) are 0 for  $n = 0$ ,  $\int_0^1 e_k(s) ds$ ,  $k = 1, 2, \dots$ , for  $n = 1$ , and 0 for  $n \geq 2$ , so that

$$B(t) = \sum_{k=1}^{\infty} \left( \int_0^t e_k(s) ds \right) H_1(\tilde{e}_k), \quad (\text{B.62})$$

where  $\{e_k(x)\}$  is an orthonormal basis in  $L^2[0, 1]$ . Figure B.1 shows ten samples of  $B(t)$  obtained from the first 600 terms of the series representation in (B.62) with  $e_k(t) = \sqrt{2} \sin((k - 1/2)\pi t)$ .  $\diamond$

*Proof* The coefficients  $a_{n_1, n_2, \dots}$  in (B.59) are zero for  $n = 0$  since they are equal to the expectation  $E[\phi(t)] = E[B(t)] = 0$ . The coefficients for  $n = 1$  are the integrals  $\int_0^t e_k(s) ds$  since, for example,

$$\begin{aligned}
 a_{1,0,0,\dots}(t) &= E[\phi(t)H_1(\tilde{e}_1)H_0(\tilde{e}_2)H_0(\tilde{e}_3)\cdots] = E[\phi(t)H_1(\tilde{e}_1)] \\
 &= E\left[\int_{[0,1]^2} 1(0 \leq u \leq t) dB(u)e_1(s) dB(s)\right] \\
 &= \int_{[0,1]^2} 1(0 \leq u \leq t)e_1(s)E[ dB(u) dB(s)] = \int_0^t e_1(s) ds,
 \end{aligned}$$

by using  $E[ dB(u) dB(s)] = \delta(s - t) ds$ . Similar calculations show that  $a_{n_1, n_2, \dots} = 0$  for  $n_1 + n_2 + \dots = n \geq 2$ . The integrals in (B.59) for  $e_k(t) = \sqrt{2} \sin((k - 1/2)\pi t)$  are  $\int_0^t e_k(s) ds = [1 - \cos((k - 1/2)\pi t)] / ((k - 1/2)\pi)$ . ▲

We conclude this section with a comment on the relationship between the Hermite polynomial expansion in (B.50) for square integrable functions in  $L^2(\mathbb{R}, \mathcal{B}, \eta)$  and the polynomial chaos expansion in (B.59) for functions in  $L^2(\Omega, \mathcal{F}^B, P)$ . The representation in (B.59) can be viewed as an infinite-dimensional analogue of the one-dimensional representation in (B.50). Truncated versions of the series in (B.59) provide approximations for the members of  $L^2(C, \mathcal{B}(C), \mu)$ , where  $C$  denotes the Banach space of real-valued continuous functions defined on  $[0, 1]$  and starting at zero,  $\mathcal{B}(C)$  denotes the Borel  $\sigma$ -field on  $C$ , and  $\mu$  is the Wiener measure. This measure can be constructed by extending the set function

$$\tilde{\mu}(A) = \int_U \prod_{i=1}^n \frac{1}{\sqrt{t_i - t_{i-1}}} \phi\left(\frac{u_i - u_{i-1}}{\sqrt{t_i - t_{i-1}}}\right) du_1 \cdots du_n \quad (\text{B.63})$$

defined on the set  $\mathcal{H}$  of cylindrical subsets  $A = \{\omega \in C : (\omega(t_1), \dots, \omega(t_n)) \in U\}$  to the  $\sigma$ -field  $\sigma(\mathcal{H})$ , where  $0 < t_1 < \dots < t_n \leq 1$ ,  $t_0 = u_0 = 0$ , and  $U \in \mathcal{B}(\mathbb{R}^n)$  ([23], Theorem 3.1.1).

### B.6.3 Multiple Wiener–Itô Integrals

We have seen in Example B.46 that the double integral  $\int_a^b \int_a^b f(s) f(t) dB(s) dB(t)$  is a polynomial chaos  $I(f)^2 - \|f\|^2$  of order  $n = 2$ . Our objectives are to define multiple integrals of the type  $I(f_1, \dots, f_k) = \int_a^b \cdots \int_a^b f_1(t_1) \cdots f_k(t_k) dB(t_1) \cdots dB(t_k)$ ,  $k = 1, 2, \dots$ , where  $f_1, \dots, f_k \in L^2[a, b]$  and  $B$  is a standard Brownian motion defined on  $(\Omega, \mathcal{F}^B, P)$ , and show that these integrals are polynomial chaoses.

The construction of  $I(f_1, \dots, f_k)$  is conceptually similar to that of the stochastic integral in Sect. 4.4. However, developments in Sect. 4.4 do not extend directly to  $I(f_1, \dots, f_k)$  since the expression of this integral for simple functions involves increments of the Brownian motion over the same time increments.

#### B.6.3.1 Two-Dimensional Wiener–Itô Integrals

Consider the stochastic integral  $I_2(f) = \int_a^b \int_a^b f(s, t) dB(s) dB(t)$ , where  $B$  denotes a standard Brownian motion and  $f \in L^2([a, b]^2)$ . We summarize the essential results in [23] (Sect. 9.2) on double Wiener–Itô integrals.

**Definition B.46** Let  $f : [a, b]^2 \rightarrow \mathbb{R}$  and call

$$\hat{f}(s, t) = (f(s, t) + f(t, s))/2 \quad (\text{B.64})$$

the symmetrization of  $f$ . Note that  $\|\hat{f}\| \leq \|f\|$ , where  $\|f\|^2 = \int_a^b \int_a^b f(s, t)^2 ds dt$  is the norm in  $L^2([a, b]^2)$ .

**Definition B.47** A function  $f : [a, b]^2 \rightarrow \mathbb{R}$  is called an off-diagonal step function if it has the expression

$$f(s, t) = \sum_{i,j=1, i \neq j}^n a_{ij} 1((s, t) \in [t_{i-1}, t_i] \times [t_{j-1}, t_j]), \quad (\text{B.65})$$

where  $a = t_0 < t_1 < \dots < t_{n-1} < t_n = b$  is a partition of  $[a, b]$  and  $a_{ij} \in \mathbb{R}$  are constants. Off-diagonal functions vanish on the diagonal  $D = \{(s, t) \in [a, b]^2 : s = t\}$  of  $[a, b]^2$  and the set of these functions is a vector space.

Consider the linear operator

$$I_2(f) = \sum_{i,j=1, i \neq j}^n a_{ij} (B(t_i) - B(t_{i-1})) (B(t_j) - B(t_{j-1})), \quad (\text{B.66})$$

defined on the set of off-diagonal functions.

**Theorem B.76** If  $f$  is an off-diagonal function, then  $I_2(f) = I_2(\hat{f})$ ,  $E[I_2(f)] = 0$ , and  $E[I_2(f)^2] = 2 \int_a^b \int_a^b \hat{f}(s, t)^2 ds dt = 2\|\hat{f}\|^2$  ([23], Lemmas 9.2.2 and 9.2.3).

**Theorem B.77** If  $f \in L^2([a, b]^2)$ , there exists a sequence  $\{f_n\}$  of off-diagonal step functions such that  $\lim_{n \rightarrow \infty} \int_a^b \int_a^b |f(s, t) - f_n(s, t)| ds dt = 0$  ([23], Lemma 9.2.4).

**Definition B.48** The m.s. limit  $I_2(f) = \lim_{n \rightarrow \infty} I_2(f_n)$ ,  $f \in L^2(\Omega, \mathcal{F}^B, P)$ , is called the double Wiener-Itô integral of  $f$ , where  $\{f_n\}$  is the sequence in Theorem B.77.

**Theorem B.78** If  $f \in L^2([a, b]^2)$ , then  $I_2(f) = I_2(\hat{f})$ , the first two moments of  $I_2(f)$  are  $E[I_2(f)] = 0$  and  $E[I_2(f)^2] = 2\|\hat{f}\|^2$ , and  $I_2(f)$  can be calculated sequentially by the formula  $\int_a^b \int_a^b f(s, t) dB(s) dB(t) = 2 \int_a^b \left[ \int_a^s \hat{f}(s, t) dB(t) \right] dB(s)$  ([23], Theorems 9.2.7 and 9.2.8).

*Example B.51* The integral  $\int_a^b \int_a^b dB(s) dB(t)$ , that is,  $I_2(f)$  for  $f = 1$ , can be calculated as the limit of integrals  $I_2(f_n)$  of simple functions (Theorem B.77) or sequentially following Theorem B.78.

Let  $a = t_0 < t_1 < \dots < t_n = b$  be a partition of  $[a, b]$  such that  $\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$  as  $n \rightarrow \infty$ . The sequences  $\{f_n\}$  and  $\{I_2(f_n)\}$  are those in (B.65) and (B.66) with  $a_{ij} = 1$ . An alternative form of  $I_2(f_n)$  is

$$I_2(f_n) = \sum_{i,j=1}^n \Delta B_i \Delta B_j - \sum_{i=1}^n (\Delta B_i)^2 = (B(b) - B(a))^2 - \sum_{i=1}^n (\Delta B_i)^2, \quad (\text{B.67})$$

where  $\Delta B_i = B(t_i) - B(t_{i-1})$ . By Theorem B.77, we have  $I_2(f) = \lim_{n \rightarrow \infty} I_2(f_n) = (B(b) - B(a))^2 - (b - a)$ , where the latter equality holds since the limit of  $\sum_{i=1}^n (\Delta B_i)^2$  is the quadratic variation of Brownian motion (Sect. 3.7.6.1).

We now calculate the same integral by the sequential rule in Theorem B.78. Since  $\hat{f} = f = 1$ , we have  $I_2(f) = \int_a^b \int_a^b dB(s) dB(t) = 2 \int_a^b \left[ \int_a^s dB(t) \right] dB(s)$ . The inner integral is  $\int_a^s dB(s) = B(s) - B(a)$  so that

$$\begin{aligned} I_2(f) &= 2 \left[ \int_a^b B(s) dB(s) - B(a) \int_a^b dB(s) \right] \\ &= 2 \left[ \int_a^b B(s) dB(s) - B(a)(B(b) - B(a)) \right]. \end{aligned} \quad (\text{B.68})$$

The integral form of Itô's formula applied to  $B(t)^2$  in the time interval  $[a, b]$  gives  $B(b)^2 - B(a)^2 = 2 \int_a^b B(t) dB(t) - \int_a^b d[B, B](t) = 2 \int_a^b B(t) dB(t) - (b - a)$  (Sect. 3.7.6.1 and Sect. 4.7), so that  $I_2(f) = (B(b) - B(a))^2 - (b - a)$  by (B.68).  $\diamond$

### B.6.3.2 Multiple Wiener–Itô Integrals

Stochastic integrals  $I_k(f) = \int_a^b \cdots \int_a^b f(t_1, \dots, t_k) dB(t_1) \cdots dB(t_k)$ ,  $k > 2$ , are constructed in the same manner as the two-dimensional Wiener–Itô integral considered in the previous section. For details, see [23] (Sect. 9.6).

As previously, let  $f(s) = \sum_{1 \leq i_1, \dots, i_k \leq n} a_{i_1, \dots, i_k} 1(s \in \times_{r=1}^n [t_{i_r-1}, t_{i_r}))$ ,  $s \in [a, b]^k$ , be a step function corresponding to a partition  $a = t_0 < t_1 < \cdots < t_n = b$  of  $[a, b]$ . An off-diagonal step function is a step function with  $a_{i_1, \dots, i_k} = 0$  if  $i_p = i_q$  for  $p \neq q$ , that is, the function is 0 whenever the intervals  $[t_{i_1-1}, t_{i_1}), \dots, [t_{i_k-1}, t_{i_k})$  are not disjoint. For an off-diagonal step function  $f$ , set  $I_k(f) = \sum_{1 \leq i_1, \dots, i_k \leq n} a_{i_1, \dots, i_k} \prod_{r=1}^k \Delta B_{i_r}$ , where  $\Delta B_{i_r} = B(t_{i_r}) - B(t_{i_r-1})$ . The symmetrization of  $f$  is the function  $\hat{f}(s_1, \dots, s_k) = (1/k!) \sum_{\pi} f(s_{\pi(1)}, \dots, s_{\pi(k)})$ , where the summation is performed over all permutations  $\pi$  of  $\{1, \dots, k\}$ , so that  $\|\hat{f}\| \leq (1/k!) \sum_{\pi} \|f\| = \|f\|$ .

Following are facts that are given without proof and parallel properties of two-dimensional integrals. If  $f$  is an off-diagonal step function, then  $E[I_k(f)] = 0$  and  $E[I_k(f)^2] = k! \int_{[a, b]^k} |\hat{f}(s_1, \dots, s_k)|^2 ds_1 \cdots ds_k$  ([23], Lemma 9.6.3).

If  $f \in L^2([a, b]^k)$ , there exists a sequence of off-diagonal step functions  $\{f_q\}$  such that  $\lim_{q \rightarrow \infty} \int_{[a, b]^k} |f(s_1, \dots, s_k) - f_q(s_1, \dots, s_k)|^2 ds_1 \cdots ds_k = 0$  and  $\{I_k(f_q)\}$  is Cauchy in  $L^2(\Omega)$ . The limit  $I_k(f) = \lim_{q \rightarrow \infty} I_k(f_q)$  in  $L^2(\Omega)$  is well defined since it does not depend on the particular sequence  $\{f_q\}$ , and is called the multiple Wiener–Itô integral of  $f$  ([23], Lemma 9.6.4). If  $f \in L^2([a, b]^k)$  and  $g \in L^2([a, b]^l)$ ,  $k \neq l$ , then  $E[I_k(f)I_l(g)] = 0$  ([23], Theorem 9.6.10). If  $f \in L^2([a, b]^n)$ ,  $k \geq 1$ , then  $I_n(f) \in K_n$ . If  $\phi \in K_n$ , there exists a unique symmetric square integrable function  $f$  defined on  $[a, b]^n$  such that  $\phi = I_n(f)$  ([23], Theorem 9.7.1), which shows that multiple Wiener–Itô integrals are polynomial chaoses.

## References

1. Hernández DB (1995) Lectures on probability and second order random fields. World Scientific, London
2. Field Jr RV, Grigoriu M (2004) On the accuracy of the polynomial chaos approximation. *Probab Eng Mech* 19(1–2):65–80
3. Grigoriu M (2006) Evaluation of Karhunen–Loève, spectral, and sampling representations for stochastic processes. *J Eng Mech ASCE* 132(2):179–189
4. Tolstov GP (1962) Fourier series. Dover Publications, Inc., New York
5. Grigoriu M (2002) Stochastic calculus. Applications in science and engineering. Birkhäuser, Boston
6. Wong E, Hajek B (1985) Stochastic processes in engineering systems. Springer, New York
7. Graf S, Luschgy H (2000) Foundations of quantization for probability distributions. Springer, New York
8. Gray R.M (1998) Quantization. *IEEE Trans Inf Theory* 44(6):2325–2383
9. Luschgy H, Pagès G (2002) Functional quantization of Gaussian processes. *J Funct Anal* 196:486–531
10. Luschgy H, Pagès G (2004) Sharp asymptotics of the functional quantization problem for Gaussian processes. *Ann Probab* 32(2):1574–1599
11. Grigoriu M (2009) Reduced order models for random functions. Applications to stochastic problems. *Appl Math Model* 33(1):161–175
12. Brabenec RL (1990) Introduction to real analysis. PWS-KENT Publishing Company, Boston
13. Bryant V (1987) Metric spaces. Iteration and applications. Cambridge University Press, Cambridge
14. Hansen VL, Functional analysis. Entering Hilbert space. Woeld Scientific, Hackensack
15. Nicolescu M (1968) Funcții Reale și Elemente de Topologie (in Romanian). Editura Didactica si Pedagogica, Bucharest, Romania
16. Naylor AW, Sell GR, Linear operator theory in engineering and science. Applied Mathematical Sciences, vol 40. Springer, New York
17. Lancaster P, Tismenetsky M (1985) The theory of matrices, 2nd edn. Academic Press, Inc., New York
18. Chen M, Chen Z, Chen G (1997) Approximate solutions of operator equations. World Scientific, New York
19. Reddy BD (1998) Introduction to functional analysis with applications to boundary value problems and finite elements. Springer, New York
20. Friedman B (1969) Lectures on applications-oriented mathematics. Wiley, New York
21. Resnick S.I (1998) A probability path. Birkhäuser, Boston
22. Schoutens W (2000) Stochastic processes and orthogonal polynomials. Lecture notes in statistics. Springer, New York
23. Hui-Hsiung Kuo, Introduction to stochastic integration. Springer, New York
24. Field RV, Grigoriu M (2007) Convergence properties of polynomial chaos approximations for  $L^2$  random variables. Technical report SAND2007-1262, Sandia National Laboratories, Albuquerque, New Mexico
25. Resnick SI (1992) Adventures in stochastic processes. Birkhäuser, Boston
26. Chung KL (1974) A course in probability theory. Academic Press, Inc., New York
27. Holden H, Øksendal B, Ubøe J, Zhang T (1996) Stochastic partial differential equations. Birkhäuser, Boston
28. Kallianpur G (1980) Stochastic filtering theory. Springer, New York

# Index

- $\sigma$ -field
  - Borel  $\sigma$ -field, 10, 11, 308, 398, 461, 464, 503, 507, 515
  - conditional expectation, 38–42
  - filtration, 43
  - generated by
    - Brownian motion, 512
    - collection of subsets, 10
    - indicator function, 18
    - random variable, 18, 20, 39, 59
    - real-valued processes, 20
    - Skorokhod topology, 308
    - topology, 16, 308
- A**
- Absorbing state, 91
- Adapted
  - $\sigma$ -field filtration, 43
  - Brownian filtration, 193
  - Brownian motion,  $97t$
  - martingale, 146
  - natural filtration, 182
  - semimartingale, 152, 156
- Adapted process, 152, 156, 158, 167, 170, 178
  - continuous, 158, 170
- Adapted stochastic process, 95, 143
  - Brownian filtration,  $97t$ , 193
  - real-valued process, 380
- Adjoint operator, 385, 498, 499
- Almost sure (a.s.) convergence, 28*f*, 29, 30, 99
  - random sequence, 135
  - refining partition, 133
- Aperiodic state, 91
- Applied SPDEs, arbitrary uncertainty, 393
  - boundary value problems
    - deterministic, 396–398
    - stochastic, 398–403
  - extended stochastic reduced order models, 427–430
  - Monte Carlo simulation, 403–412
  - stochastic collocation method, 440–446
  - stochastic Galerkin method, 430
    - physical space discretization, 436
    - probability space discretization, 431–435
    - solution existence and properties, 436–440
  - stochastic reduced order model method, 411–413
  - stochastic reduced order models, 413–416
    - boundary value problems, 420–427
    - initial-boundary value problems, 416–420
- Applied SPDEs, small uncertainty, 446
  - Neumann series, 449–450
  - perturbation series, 448–449
  - Taylor series, 447–448
- a.s. convergence. *See* Almost sure (a.s.) convergence
- B**
- Babuška–Lax–Milgram theorem, 272
- Banach fixed point theorem, 483
- Banach space, 485, 486
- Basis function, 228, 230, 232
  - finite element, 187, 192
- Bayes’ formula, 13, 259
- Bernoulli random variables, 22
- Bernstein polynomial, 365, 366
- Bessel’s equation, 493



**B** (*cont.*)

Bessel's inequality, 493, 494, 500  
 Beta distribution, 224, 354, 404  
 Beta random variable, 407  
 Beta standard variable, 442  
 Bilinear form, 340, 359, 397  
   bounded, 400, 496  
   continuous, 398, 436  
   elliptical, 397, 398, 437, 496  
 Bilinear functional, 272, 273, 275, 339  
 Bochner's theorem, 70–74, 78  
 Borel  $\sigma$ -field, 11, 398, 503, 509, 515  
   generated by  
     open balls, 308, 464  
     stochastic process, 468  
     topology, 461  
 Borel–Cantelli lemma, 21, 22  
 Borel measurable function, 20  
 Borel measurable mapping, 49, 461, 468  
 Borel measure, 11  
 Borel sets, 11, 63, 70  
 Boundary conditions  
   deterministic, 387  
   Dirichlet, 182, 439  
   essential, 396  
   homogeneous, 396, 448, 449  
   natural, 253  
   Neumann, 182  
 Boundary value problem, 420  
   deterministic, 396  
   Dirichlet ( $q = 0$ ), 184–188  
   Dirichlet ( $q \neq 0$ ), 188–190  
   mixed, 190–192  
   stochastic, 221, 393, 394, 398–403, 427, 436, 446  
 Bounded convergence, 29  
 Bounded linear operator, 488, 489, 498, 499, 501  
 Bounded operator, 498  
 Bounded variables, 210–211  
 Bounded variation, 79, 80, 81, 130, 131, 176  
 Bounded linear operators, 498–500  
 Brownian filtration, 140, 193  
 Brownian motion, 62  
   as a martingale, 138, 140, 169, 176  
   continuous time martingales, 97–99  
   geometric Brownian motion, 163, 166, 170, 173, 253, 255, 305  
   non-differentiable sample paths, 4  
   quadratic variation, 98, 100, 105, 106f, 148, 516  
   reflected at two thresholds, 182, 183  
   reflected at zero

local time, 181

  Tanaka's formula, 181–184

unbounded variation, 98

Brownian motion integrators

  stochastic integral, 136

$\mathcal{H}_0^2$ , integrands, 137–138

$\mathcal{H}^2$ , integrands, 138–142

**C**

Cauchy–Schwarz inequality, 41, 46, 68, 298, 340, 341, 385, 398, 400, 406, 503

  Hilbert spaces, 490, 491

  linear functional, 496, 497

  metric space, 476

  normed-linear spaces, 484–486

  normed spaces, 504

  second moment calculus, 76

Cauchy sequence, 395, 399, 482

  in Banach space, 489

  in  $L^2$ , 78, 138, 145

  in metric spaces, 487

  m.s. Riemann–Stieltjes integrals, 79, 83

  in probability, 29

Central limit theorem, 29

Central moment of order, 24

Change of measure

  absolute continuity of measures, 32

  continuity of probability measure, 21

  density function, 32–36

  Girsanov's theorem, 193–196

  Radon–Nikodym theorem, 31

Chapman–Kolmogorov equation, 90, 93

Characteristic function, 94

  integro-differential equation, 314

  Lévy–Khintchine formula, 104, 105

  Liouville equation, 291

  random variables, 36–38

Chebyshev's inequality, 137, 229,

  309, 348

  m.s. convergence, 28

Chebyshev polynomials, 235

  modified, 232, 428

  products of, 428

Closed set, 479–481

Coefficient of variation, 24

Collocation method, 224, 267, 276

Colored noise, 4, 109, 175, 177, 246

Compact operator, 499, 502

Compact sets, 479–481

Compact support, 188, 395

Compensated Poisson process, 100, 153

  compound process, 101, 142

- Complete metric space, 308, 480, 483, 487
  - Completeness, 143
  - Complete sets, 480
  - Completion theorem, 487
  - Compound Poisson process, 62*f*, 65, 69, 78, 79, 93, 94, 99–103
    - and  $\alpha$ -stable processes, 105
    - and Brownian motion, 151, 281, 310
    - covariance function, 108
    - semimartingales, 146
  - Conditional density function, 34
  - Conditional distribution function, 34
  - Conditional expectation
    - change of fields, 38
    - conditional distribution, 39
    - conditional probability, 34–40
    - defining relation, 40, 42
    - probability with respect to  $\sigma$ -field, 35–40
  - Conditional probability, 13, 34
  - Conjugate priors, 206
  - Consistency condition, 64, 390
  - Continuous in mean square, 76–77
  - Continuous martingale, 140, 143
  - Continuous stochastic process, 163
  - Continuous time martingales, 95–97
    - Brownian motion process, 97–99
    - Lévy processes, 103–107
    - Poisson processes
      - and compound Poisson processes, 99–103
      - white noise processes, 107–109
  - Continuous time linear systems, 241
  - Continuous time nonlinear systems, 248
  - Continuum eigenfunction, 501
  - Continuum eigenvalue, 501
  - Contraction, in metric space, 482, 483
  - Contraction factor, 482
  - Convergence
    - distribution, 28
    - $L^p$ , 142, 445–446
    - polynomial chaos representations, 361
    - probability, 28, 145, 159
    - probability one, 76, 89
    - strong law of large numbers, 29
    - weak law of large numbers, 29
  - Convergence modes
    - almost sure (a.s.) convergence, 28, 28*f*, 271
    - bounded convergence, 29, 264
    - central limit theorem, 29
    - coordinate convergence, 482
    - dominated convergence, 29, 41, 142, 508
    - mean square (m.s.) convergence, 28, 101, 136, 404
    - Cauchy–Schwarz inequality, 406
    - finite difference scheme, 392
    - Itô integral, 133–134
    - real-valued random function, 77
    - monotone convergence, 29, 41
    - monotone convergence theorem, 29, 33
    - Skorokhod convergence, 308
    - weak convergence, 272, 308, 497–498
  - Correlation, 68, 78, 381
  - Countable sample space, 14
  - Covariance, 25, 38, 86
  - Covariance function, 67, 69, 115, 124*f*, 125*f*, 414*f*, 432
    - in translation space, 119*f*
    - target covariance function, 118, 120, 321
  - Covariance matrix, 25, 212
- D**
- Decomposition method
    - Cholesky decomposition, 48, 339, 458
    - Doob's decomposition, 44, 45, 143
    - Doob–Meyer decomposition, 143–144, 145, 149
    - Lévy decomposition theorem, 104
    - Lévy–Khintchine formula, 104
    - unique, 486, 492
  - Degrading systems, 327
  - Dense subspace, 487, 501
  - Density (probability density) function, 32–36
    - conditional random variable, 357
    - Fourier pairs, 37
    - Gaussian (normal), 33
    - standard bivariate Gaussian vector, 35
    - conditional, 34
  - DEs. *See* Deterministic equations (DEs)
  - Deterministic equations (DEs), 1, 3, 317, 361, 402
  - Differential equations
    - characteristic function, 172, 173
    - density, Fokker–Planck equation, 251, 253, 288, 313, 316
    - moments, 172
  - Diffusion coefficient, 167
  - Diffusion process, 167
  - Directional wind speed, for hurricanes, 211–213
  - Dirichlet boundary value problem ( $q = 0$ ), 184–188
  - Dirichlet boundary value problem ( $q \neq 0$ ), 188–190
  - Dirichlet problem, 188
  - Discrete random variable

**D (cont.)**

- $\sigma$ -field generated by, 39, 40
- conditional expectation given by, 38–42
- Discrete time linear systems, 239
- Distance, 476
- Distribution function, 32–36
  - Gaussian, 34–35, 48, 214, 408
  - non-Gaussian, 36, 48–50, 238
  - multivariate translation distribution, 50
- Dominated convergence, 41
- Dominated convergence theorem, 29
- Doob–Dynkin lemma, 143, 184, 252, 437
- Doob maximal  $L^2$  inequality, 47
- Doob–Meyer decomposition, 143
- Doob’s decomposition, 44
- Double Wiener–Itô integral, 246, 515, 516
- Drift, 167
- Drift coefficient, 167, 490
- Dual space, 490
- Dynkin formula, 186

**E**

- Eigenfunction, 83
  - correlation function, 223, 463
  - Karhunen–Loève representation, 85
- Eigenvector, 354, 502
- Eigenvalue, 79, 83, 349
  - correlation function, 223, 463
  - differential equation, 383
  - distribution, 352f
  - Karhunen–Loève representation, 85
- Elliptical partial differential operator, 396
- Equivalent linearization method, 374–375
- Ergodic, 88
- ESROMs. *See* Extended stochastic reduced order models (ESROMs)
- Essentially bounded, 505
- Estimator, 50
  - in Monte Carlo simulation, 50–53
- Euclidean norm, 169
- Euclidean space, 476
- Existence and uniqueness solutions
  - Galerkin solutions, 432
  - stochastic differential equations, 168
    - Lax–Milgram theorem, 340
    - strong solutions, 276
    - weak solutions, 291, 359
  - stochastic integral equation, 168, 178
  - stochastic partial differential equations, 381
  - stochastic problems, 433
- Expectation operator, 22
- Exponential process, 193

- Extended stochastic reduced order models (ESROMs), 350, 427, 469, 470

**F**

- Fatou’s lemma, 30
- Feynman–Kac formula, 189
- Filtered probability space, 43
- Filtration
  - adapted stochastic process, 95, 143
  - Brownian, 140, 193
  - natural, 43
  - right continuous, 33
  - standard Brownian, 135, 140, 193, 382
- Finite difference method, 388
- Finite dimensional distributions, 17, 62–64
- Finite element solution, 187, 192, 239, 380, 391
- Fixed point theorem. *See* Banach fixed point theorem
- Fokker–Planck equation, 251, 288
- Fourier coefficients, 492
- Fourier series, 120, 218
- Fourier transform, 122, 251, 289
- Fredholm integral equation, 450
- Fubini’s theorem, 22, 25, 26, 27, 54
  - expectation element, 180, 313
  - independence, 168
  - integration order
    - of physical and probability spaces, 246
  - in measurable random function, 60, 80, 172, 399

**G**

- Galerkin method, 290, 357, 430, 431, 438–441
- Gamma density function, 53
  - normal, 205, 259
- Gamma distribution, 349, 468
- Gamma function, 53
- Gaussian measure, 510
- Gaussian (normal) distribution, 216
  - density, 33, 34–35
  - linear transformations, 37
  - multivariate, 49–50
  - standard normal, 358
- Gaussian process
  - Brownian motion, 168
  - colored noise, 175
  - Ornstein–Uhlenbeck process, 177
  - stationary, 109–111
  - white noise, 108, 109
- Gaussian random variable. *See* Gaussian (normal) distribution

- Gaussian variables, 203
  - Bayesian method, 205–206
  - frequentist method, 204–205
- Gaussian white noise, 166–177
- Geometric Brownian motion, 163, 166, 253
  - Itô equation, 177
  - stochastic differential equation, 282
- Girsanov theorem, 193–196
- Gram–Schmidt orthogonalization procedure, 493, 502
- Green’s theorem, 397
- Gronwall’s inequality, 319, 385, 451
  
- H**
- Hausdorff space, 478
- Hermite polynomial, 360, 362, 507–510
- Hermitian, 72
- Hilbert space, 490, 491
  - basis and Fourier representations, 491–495
  - bounded linear operators, 498–501
  - linear functionals, 496–497
  - spectral theory, 501
  - weak convergence, 497
- Holder’s inequality {Hölder}, 503, 504
- Homogeneous chaos, 358, 510–511
  - expansion, 511
- Homogeneous Poisson process, 142, 211
- Homeomorphic spaces, 478
  
- I**
- Inclusion–exclusion formula, 12
- Independence
  - $\sigma$ -field, 10, 19, 20
  - events, 10, 18–21
  - Gaussian field, 224
  - probability space, 19
  - random variables, 18, 313
- Inequalities
  - Cauchy–Schwarz, 76
  - Chebyshev, 28, 229
  - Doob, 47
  - Hölder, 503, 504
  - Jensen, 23, 41
  - Minkowski, 504
- Infinitely divisible characteristic function
  - $\alpha$ -stable, 105, 310
  - Lévy–Khintchine formula, 104–105
  - properties, 9, 23, 24
- Infinitesimal generator, 166
- Inner product, 490, 491, 496, 506
- Integrals of random variables
  - Fatou’s lemma, 30
  - Lebesgue’s theorem, 30
  - properties, 36, 37
- Isometric mapping, 478
- Isometric operator, 271
- Isometry, 137, 138
- Itô calculus, 305
  - arbitrary semimartingales, 158
  - continuous semimartingales, 156
- Itô formula
  - multi-dimensional, 512
  - one-dimensional, 52, 389
  - $\mathbb{R}$ -valued semimartingales, 155
    - arbitrary semimartingales, 158–162
    - continuous semimartingales, 156–158
  - $\mathbb{R}^d$ -valued semimartingales, 162–164
  - and Stratonovich integrals, 164–165
- Itô calculus, applications, 165
  - Girsanov’s theorem, 193–196
  - random walk method, 184
    - Dirichlet boundary value problem ( $q = 0$ ), 184–188
    - Dirichlet boundary value problem ( $q \neq 0$ ), 188–190
    - mixed boundary value problem, 190–192
  - stochastic differential equations, 166
    - Gaussian white noise, 166–177
    - semimartingale white noise, 178–181
  - Tanaka’s formula, 181–184
- Itô integral, 132, 135, 145, 164
- See also* Stochastic integral
- Itô isometry, 137, 138
- Itô stochastic differential equation, 165, 166
  
- J**
- Jensen’s inequality, 23, 41, 96
  - for random variables, 95
- Joint density function, 223
- Joint distribution, 17, 59
- Joint probability density function, 223
  
- K**
- Karhunen–Loève expansion, 84–86, 455–459
- Kolmogorov’s continuity criterion, 64, 66–67
  
- L**
- $L^2$  ( $\Omega$ ,  $F$ ,  $P$ ), 41, 65–68, 138, 339, 340, 358, 359
- $L^p$  ( $\Omega$ ,  $F$ ,  $P$ ), 30, 43
- See also* Probability space
- $L^p$  spaces, 503, 504

**L (cont.)**

- as normed spaces, 504–505
- useful inequalities, 503
- Laplace operator, 186
- Law of total probability, 13
- Lax–Milgram theorem, 340, 401, 497
- Lebesgue integral, 490
- Lebesgue measure, 11, 26, 81, 437
  - deterministic BVPs, 401
  - homogeneous chaos, 511
  - random fields, 112
  - real-valued functions, 485
  - spectral density, 71
- Lebesgue's theorem, 30
- Legendre polynomials, 291, 293, 362
- Lévy measure, 104, 105
- Lévy process
  - Lévy decomposition, 104
  - Lévy–Khintchine formula, 104–105
  - quadratic variation, 106f
- Linear functionals, 498–499
- Linearity, 41, 486, 492
- Linear operator, 488, 497–498
- Linear spaces, 483–484
- See also* Vector spaces
- Linear stochastic differential equation, 271
- Linear transformation, 37
- Lipschitz condition, 169
- Localizing sequence, 96
- Local martingale, 96, 147
- Local solution
  - Feynman–Kac functional, 189
  - random walk method, 184
  - Schrödinger equation, 183
- Local time, 181
- Lyapunov exponent, 306

**M**

- Marginal density function, 63, 92, 249, 283, 284, 317f
  - homogeneous random field, 433
- Marginal distribution, 63, 87, 265, 468, 471f
- Markov process, stationary, 92–93, 129
- Markov property, 167
- Markov random function, 90
  - Markov chains, 90–92
    - Chapman–Kolmogorov equation, 93
  - Markov process, 92–93
- Martingale
  - continuous time, 95, 97
  - discrete time, 42–45, 95
  - Doob decompositions, 44
  - Doob inequality, 47
  - Doob–Meyer decomposition, 143–145
    - as integrators, 142, 144, 146
  - Jensen inequality, 23, 41, 95
    - stopped, 44, 96
  - submartingale, 44, 96
  - supermartingale, 95, 96
    - variation and covariation, 148–152
- Mean square (m.s.) convergence, 28, 384, 390, 392, 394
- Mean square error, 38, 374, 455
- Measurable function, 16–18, 25, 60
- Measurable sets, 11, 39, 65
- Measurable space, 10, 11, 16–18
  - probability space, 43, 50, 51, 63
  - Radon–Nikodym derivatives, 31, 42
- Metric space, 475–478, 487
  - closed set, 479–481
  - compact sets, 479–481
  - complete set, 479–481
  - completion of, 487
  - contraction, 482
  - sequences, 481–482
  - topology generated by, 477
- Microstructures, probabilistic models for, 222
  - linear models
    - dependent coefficients, 227–234
    - independent coefficients, 223–227
- Minkowski inequality, 504
- Mixed boundary value problem, 190–192
- Modulus inequality, 41
- Moment of order, 24, 50, 173, 306, 320, 325–327, 345, 351, 469
  - marginal, 422
- Monotone convergence, 41
- Monotone convergence theorem, 29, 33
- Monotonicity, 41
- Monte Carlo simulation, 337, 345
  - improved measure change, 50
- Fourier series, 120
  - linear differential equations, 286, 342
  - memoryless transformations, 48, 171
  - non-Gaussian process and field, 113, 118f, 238
  - non-stationary Gaussian process and field, 109, 119, 124, 230
  - random variable, 203
  - stationary Gaussian process and field, 103, 123
    - random fields, 112–113
    - sampling theorem, 459, 460
    - spectral representation, 81
    - stochastic processes, 109–111

translation vector processes, 113–119  
 m.s. convergence. *See* mean square  
   convergence  
 m.s. Riemann–Stieltjes integrals, 79  
 Multiple Wiener–Itô integral, 515–517  
 Multivariate Gaussian, 50  
 Multivariate translation distribution, 50

## N

Natural boundary, 253  
 Neumann series, 300–303  
 Neumann series method, 373, 449–450  
 Noise induced transition, 311  
 Noninformative density, 206  
*See also* Vague prior density  
 Norm  
   Euclidean, 485, 489  
   operator, 488  
 Normed linear spaces, 484  
   basis and separability, 487  
   metric spaces, completion of, 487  
   operators, 488, 490  
 Normed vector space, 485, 487  
 Null set, 11, 380  
 Nyquist sampling rate, 460

## O

ODE. *See* Ordinary differential equation  
   (ODE)  
 Off-diagonal step function, 516–517  
 Open ball, 16, 242, 308, 477, 478  
 Open set, 59, 60, 477, 479, 398  
 Operator  
   bounded, 488, 489, 497, 498, 499, 502  
   closed, 480, 481, 492  
   compact, 499, 500, 502  
   continuous, 489  
   differential, 192, 395, 396  
   isometric, 487  
   linear, 41, 81, 83, 491, 494, 500, 502–504,  
     506, 516  
   matrix, 500, 502  
   projection, 496, 499  
   self-adjoint, 83, 498, 499, 502, 506  
 Operator norm  
   bounded operator, 498  
   self-adjoint operator, 502, 506  
 Optional stopping theorem, 47  
 Ordinary differential equation (ODE),  
   172–173, 180, 287  
   finite differences, 387, 390, 394

linear, 2, 5, 172–173, 215, 222, 245, 251,  
   255, 273–274, 390, 394  
   second order, 237  
   stochastic, 3  
   white noise, 5  
 Ornstein–Uhlenbeck process, 168, 171–173,  
   177, 255, 260, 277, 284  
 Orthogonal functions, 40, 495, 508–509  
 Orthogonal increments, 81–83, 457  
 Orthogonal polynomials, 506, 507  
   Hermite polynomials, 513  
   homogeneous chaos, 510–511  
   multiple Wiener–Itô integrals, 517  
   two-dimensional Wiener–Itô integrals,  
     515–517  
 Orthogonal projection, 40  
 Orthonormal basis, 359, 493, 495, 500, 502,  
   508, 512  
 Orthonormal sequence, 492, 498, 502, 508  
 Orthonormal set, 494

## P

Parallelogram law, 490, 491  
 Parametric models, 464  
   Karhunen–Loève expansion, 457, 459, 463  
   sampling theorem, 459, 460  
   spectral representation, 457, 458, 459  
 Parseval’s identity, 495  
 Perturbation method, 269, 372  
 Perturbation series, 368, 371, 448  
 p-integrable martingale, 43  
*See also* Submartingale; Supermartingale  
 Poincaré–Friedrichs inequality, 398, 402  
 Poisson process  
   compensated, 153  
   and compound Poisson process, 99–103  
   quadratic variation, 106f, 148  
   filtration, 95, 99  
 Polynomial chaos, 362–376, 436  
 Positive definite, 88, 475, 484, 490  
 Posterior densities, 206, 209f, 214, 259  
 Predictable process, 45, 96, 143, 144  
 Predictable stochastic processes, 44, 96  
 Probability density function, 32  
 Probability distribution function, 17  
 Probabilistic models, 201  
   random functions, 214, 230  
   multi-phase materials, 217  
   probabilistic models for microstruc-  
     tures, 221, 379  
   uncertain parameters, systems with,  
     215, 216

**P (cont.)**

- random variables, 203
  - bounded variables, 210
  - directional wind speed for hurricanes, 211
  - Gaussian variables, 203, 211
  - translation variables, 206
- Probability measure, 11
  - extension of, 15–16
- Probability space, 9–16
  - $\sigma$ -field, 10–11
  - construction of, 14
    - countable sample space, 14
    - product probability space, 14–15
  - sample space, 9–10
- Probability theory, essentials of
  - characteristic functions, 36–38
  - conditional expectation, 38–42
  - density functions, 32–36
  - discrete time martingales, 42–47
  - distribution functions, 32–36
  - expectation operator, 22–27
  - independent events, 18–21
  - measurable functions, 16–18
  - Monte Carlo simulation, 47–53
    - estimators, 50–53
    - Gaussian variables, 48
    - non-Gaussian variables, 48–50
  - probability measure, 11–14
    - extension, 15–16
  - probability space, 9–11
    - countable sample space, 14
    - product probability space, 14–15
  - Radon–Nikodym derivative, 31–32
  - random elements, 16–18
  - random variables
    - convergence of sequences, 27–30
    - sequence of events, 21–22
- Product probability space, 14–15, 339
- Product sample space, 14
- Pythagoras' theorem, 492, 494, 495

**Q**

- $q$ -dimensional Euclidean space, 478
- Quadratic variation, 148, 106
  - Brownian motion, 98, 518
  - compensated Poisson process, 100, 153
- Quadratic variation process, 100, 105, 126, 171
- Quadratic variation and covariation
  - integration by parts, 150
  - polarization identity, 152
- Quantizers, 463–466

**R**

- Radon–Nikodym derivative, 31–32, 42, 51
- See also* Monte Carlo simulation algorithms; Spectral density
- Random field, 73, 79, 112, 220, 224, 436
- Random function
  - finite dimensional distribution, 62–64
  - Monte Carlo simulation, 109
    - Gaussian stationary random functions, 109–113
    - non-stationary Gaussian processes, 119–124
    - translation vector processes, 113–119
  - multi-phase materials, 219, 221
    - data set, 219–222
    - spherical harmonics, 218–219
  - probabilistic models for microstructures, 221
    - dependent coefficients, 227, 235
    - independent coefficients, 223, 235
  - sample properties, 64–67
  - second moment calculus, 76
    - integrals, 79–80
    - Karhunen–Loève expansion, 83–85
    - mean square continuous, 76–77
    - mean square differentiable, 77–79
    - spectral representation, 81–83
  - second moment properties, 67–70
  - stochastic processes, classes of, 85
  - continuous time martingales, 95–109
  - ergodic random functions, 88–90
  - Gaussian random functions, 86
  - independent increments, 93–95
  - Markov random function, 90–93
  - translation random functions, 86–88
  - uncertain parameters, 215–217
  - weakly stationary random functions, 70
    - $\mathbb{R}$ -valued stochastic processes, 72–73
    - $\mathbb{R}$ -valued random variables, 73–75
    - $\mathbb{R}$ -valued random variables, 70–72
- Random variable
  - arbitrary, 22, 23, 34, 40, 85
  - characteristic function, 36–38, 94–95, 101, 108, 136, 225
  - continuous, 16, 64–65, 77, 96
  - convergence, sequences of, 28f
  - density, 34, 93, 203, 288, 466
  - discrete, 39
  - distribution, 17, 32, 36, 48, 56, 260, 408, 467, 473
  - expectation, 22–23, 25, 36, 38–42, 51, 285, 307, 347
  - moment, 24, 67–70
  - P-integrable, 401

- standard deviation, 17, 220
- variance, 76
- Random vector
  - characteristic function, 265, 269, 287, 313
  - correlation, 87, 325–332
  - covariance matrix, 38, 49, 331
  - density, 35, 94
  - distribution, 17, 62–63
  - distribution function, 234
  - expectation, 22, 25
  - Gaussian, 34–35, 37, 49, 97, 438
  - independence, 18, 313
  - joint density function, 207–208
  - joint distribution function, 17
  - moments, 24–25, 344
  - second moment properties, 67–70
- Random vectors, 25, 295, 341, 414
- Random walk method, 184
  - Dirichlet boundary value problem ( $q = 0$ ), 184–188
  - Dirichlet boundary value problem ( $q \neq 0$ ), 188–190
  - mixed boundary value problem, 190–192
- Reliability method, 367, 368
- Residual spectrum, 501
- Resolvent, 501
- Riemann integral, 163, 175
- Riemann–Stieltjes integral, 79, 130–131, 143
- Riesz representation theorem, 498–499
- Right continuous filtration, 55
  
- S**
- SAEs. *See* Stochastic algebraic equations (SAEs)
- Sample space, 9–10
  - countable, 14
- Schauder basis, 487
- Schrödinger equation, 182, 184
- SDEs. *See* Stochastic differential equations (SDEs)
- Second moment calculus for processes
  - expectation and mean square integrals, 76
  - mean square continuity, 77
  - mean square differentiation, 77–79
  - mean square integration, 83
  - spectral representation, 81, 109
- Second moment properties, 67–70
  - stochastic process, 238, 241, 246, 248
- Self-adjoint operator, 83
- Semimartingale, 155–164
- Semimartingale white noise, 178–181
- Separability
  - by a basis, 487
  - topological space, 487
- Separable stochastic process, 66–67
- Separation of variables, 384
- Sequence of events, 21–22
- Sequences, 481, 482
- SEs. *See* Stochastic equations (SEs)
- Simple stochastic process, 60, 469
- Smolyak formula, 367
- Solution of stochastic differential equation, 168, 301
- Space
  - Banach, 487
  - complete metric, 480
  - Euclidean, 414, 485
  - Hilbert, 490
- Spectral density function, 71, 82
- Spectral distribution, 71–74, 89, 457
- Spectral theory, 501
- Stochastic partial differential equations (SPDEs)
- SROMs. *See* Stochastic reduced order models (SROMs)
- Standard normal distribution, 225
- Standard normal random variable, 225
- State augmentation, 286, 287
- State space, 90–92
- Stationary distribution, 92
- Stationary increments, 97, 99, 103
- Stationary process
  - in strong sense, 167, 170
  - in weak sense, 167, 272
- Stochastic algebraic equations (SAEs), 346
  - arbitrary uncertainty, 337
  - general considerations, 339
  - Monte Carlo method, 339, 342
  - reliability method, 368
  - stochastic collocation method, 363
  - stochastic Galerkin method, 359, 361
  - stochastic reduced order model method, 344–357
- small uncertainty, 370
  - equivalent linearization, 374–375
  - Neumann series, 374, 375
  - perturbation series, 371, 372
  - Taylor series, 369, 371
- Stochastic collocation method, 295, 299, 363, 440–446
- Stochastic difference equations, 254
  - conditional analysis, 260, 276
  - Monte Carlo simulation, 260, 263
  - perturbation series, 269
  - stochastic Galerkin and collocation methods, 267



**S** (*cont.*)

- stochastic reduced order models, 265, 276
  - Taylor series, 267–269
- Stochastic differential equations (SDEs), 166, 271
  - additive noise, 263
  - Brownian motion input
    - diffusion process, 313, 329
  - conditional analysis, 277
  - conditional Monte Carlo simulation, 278, 279
  - definition, 166, 169
  - existence and uniqueness of solution, 179
  - geometric Brownian motion, 170, 255, 305
  - Gaussian white noise, 166–177
  - linear, 251
  - Monte Carlo simulation, 276, 277
  - multiplicative noise, 237, 255
  - Neumann series, 300, 303
  - perturbation, 300, 303
  - semimartingale input, 5, 158
    - white noise, 178–181
  - solution, 317
  - state augmentation, 287
  - stochastic collocation method, 295, 299
  - stochastic Galerkin method, 290, 294
  - stochastic reduced order models, 289
  - Stratonovich integral, 164–166
  - Taylor series, 300–301
  - Wong–Zakai theorem, 175
- Stochastic equations (SEs), 1
  - deterministic coefficients
    - discrete time linear systems, 239
    - continuous time linear systems, 241, 248
    - continuous time nonlinear systems, 248, 251
- Stochastic Galerkin method, 290–295, 351–363
  - factors affecting accuracy, 362
- Stochastic integral
  - $\int B dB$  and  $\int N dN$ , 131–136
  - associativity, 147
  - Brownian motion integrators, 136
    - $\mathcal{H}_0^2$  integrands in, 137, 138
    - $\mathcal{H}^2$  integrands in, 138–142
  - Martingale integrators, 142–146
  - Itô integral, 137
  - preservation, 147
  - quadratic variation and covariation processes, 148–152
  - Riemann–Stieltjes integrals, 130–131
  - semimartingale, 146, 148
    - simple predictable integrand, 137–142
    - Stratonovich integral, 132, 164
- Stochastic integral equation, 166
- Stochastic partial differential equations (SPDEs), 5, 379, 380–386
  - discrete approximations of, 386–392
- Stochastic process
  - adapted, 178
  - classes of, 85
  - continuous time, 95
  - correlation, 70
  - covariance, 69, 70, 74
  - discrete time, 42, 95, 254
  - distribution, 32
  - expectation function, 22
  - finite dimensional distributions, 62–64
  - Gaussian, 86
  - measurable, 26
  - Poisson, 142, 151
  - sample properties, 230
  - second moment properties, 216, 227
- Stochastic process increment
  - Brownian motion, 93, 98, 168, 307
  - independent increments, 62, 69, 93, 107
    - Gaussian, 4, 97, 97t, 164
  - Markov process, 167
  - orthogonal increments, 81–83, 457
  - Poisson process, 69–70, 93
    - characteristic function, 94
  - stationary increments, 97
  - stationary, independent increments, 93
- Stochastic reduced order models (SROMs), 289, 290, 317–320, 393, 464–469
  - Borel measurable mappings, 468
  - extended stochastic reduced order models (ESROMs), 350–357
  - Gamma random variable, 466
  - uncertain dynamic systems, 317–320
- Stochastic stability, applications of, 304–310
- Stopping time, 46
- Stratonovich integral, 132, 133
  - and Itô formula, 164–165
- Stratonovich stochastic differential equation, 165
  - equivalent Itô stochastic differential equation, 175, 197
- Strong continuity, 76
- Strong convergence, 360
- Strong law of large numbers (*see also* under Convergence), 29
- Strong Markov property, 167
- Strong solution (of stochastic differential equation), 167

Strong uniqueness (of stochastic differential equation), 167  
 Sturm–Liouville equation, 506  
 Submartingale, 43  
 Supermartingale, 43  
 Sup-metric, 477  
 Symmetry, 475

## T

Tanaka's formula, 166, 181–184  
 Target covariance function, 118, 120, 321  
 Taylor series, 300–301  
 Taylor series method, 371, 374, 450  
 Time change  
   of local martingale, 96  
   simplest case, 240  
 Time change formula Itô integrals, 164  
 Total variation process, 148, 150  
 Triangle inequality, 475, 484  
 Transition density, 92  
 Transition probability matrix, 90–92  
 Translation process, 115, 118, 468  
 Translation variables, 206  
   Bayesian method, 208–209  
   frequentist method, 207–208

## U

Unbounded operator, 501  
 Uniform distribution, 216  
 Uniform integrability, 83

## V

Vague prior density, 206  
*See also* Noninformative density  
 Vector spaces, 483  
   finite dimensional, 484, 485  
   infinite dimensional, 485  
    $n$ -dimensional, 484, 486  
   normed, 485, 487  
 Version, 2, 6, 66, 277

## W

Weak convergence, 308, 497  
 Weak law of large numbers (*see also* under Convergence), 29  
 Weak solution, 272, 339  
 Weak uniqueness, 167  
 Weierstrass theorem, 366  
 Well posed  
   boundary value problem, 396  
   linear partial differential equations, 390  
   SDPE, 386, 396  
 White noise processes, 3–4, 79, 107–109, 174–175  
 Wiener–Askey polynomials, 362  
 Wiener–Itô integrals  
   multiple, 517  
   two dimensional, 515–517  
 Wiener measure, 515  
 Wong–Zakai theorem, 175